11-я РОССИЙСКАЯ МУЛЬТИКОНФЕРЕНЦИЯ ПО ПРОБЛЕМАМ УПРАВЛЕНИЯ

ПРЕДСЕДАТЕЛЬ ПРЕЗИДИУМА МУЛЬТИКОНФЕРЕНЦИИ
АКАДЕМИК РАН В. Г. ПЕШЕХОНОВ



2–4 октября 2018 г.
ГНЦ РФ АО «КОНЦЕРН «ЦНИИ «ЭЛЕКТРОПРИБОР»
Санкт-Петербург

ПРИ ПОДДЕРЖКЕ:

- РОССИЙСКОГО ФОНДА ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ (РФФИ)
- ОТДЕЛЕНИЯ ЭНЕРГЕТИКИ, МАШИНОСТРОЕНИЯ, МЕХАНИКИ И ПРОЦЕССОВ УПРАВЛЕНИЯ РАН
- ОТДЕЛЕНИЯ НАНОТЕХНОЛОГИЙ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ РАН
- МЕЖДУНАРОДНОЙ ОБЩЕСТВЕННОЙ ОРГАНИЗАЦИИ «АКАДЕМИЯ НАВИГАЦИИ И УПРАВЛЕНИЯ ДВИЖЕНИЕМ»
- НАУЧНОГО СОВЕТА РАН ПО ТЕОРИИ УПРАВЛЯЕМЫХ ПРОЦЕССОВ И АВТОМАТИЗАЦИИ
- ОБЪЕДИНЕННОГО НАУЧНОГО СОВЕТА ПО КОМПЛЕКСНОЙ ПРОБЛЕМЕ «ПРОЦЕССЫ УПРАВЛЕНИЯ И АВТОМАТИЗАЦИЯ» РАН
- САНКТ-ПЕТЕРБУРГСКОЙ ТЕРРИТОРИАЛЬНОЙ ГРУППЫ РОССИЙСКОГО НАЦИОНАЛЬНОГО КОМИТЕТА ПО АВТОМАТИЧЕСКОМУ УПРАВЛЕНИЮ
- ЖУРНАЛА «ГИРОСКОПИЯ И НАВИГАЦИЯ»

МАТЕРИАЛЫ КОНФЕРЕНЦИИ «ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В УПРАВЛЕНИИ» (ИТУ-2018)

Санкт-Петербург
2018

УДК 681.5

Материалы конференции «Информационные технологии в управлении» (ИТУ-2018).
СПб.: АО «Концерн «ЦНИИ «Электроприбор», 2018. 628 с.

В настоящем сборнике представлены тексты докладов, состоявшихся на конференции «Информационные технологии в управлении» (ИТУ-2018).

Тексты докладов публикуются в авторской редакции.

ОРГАНИЗАТОРЫ КОНФЕРЕНЦИИ:

- ГОСУДАРСТВЕННЫЙ НАУЧНЫЙ ЦЕНТР РФ АО «КОНЦЕРН «ЦНИИ «ЭЛЕКТРОПРИБОР»
- ФГБУН «САНКТ-ПЕТЕРБУРГСКИЙ ИНСТИТУТ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ РАН»
- ФГАОУ ВО «САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ»
- ФГАОУ ВО САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ «ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
- НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ МНОГОПРОЦЕССОРНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ ИМ. АКАДЕМИКА А.В. КАЛЯЕВА ЮЖНОГО ФЕДЕРАЛЬНОГО УНИВЕРСИТЕТА
- АО «РКЦ «ПРОГРЕСС»
- ПАО «РАКЕТНО-КОСМИЧЕСКАЯ КОРПОРАЦИЯ «ЭНЕРГИЯ» ИМЕНИ С.П. КОРОЛЁВА»
- ФГБУН «ИНСТИТУТ ПРОБЛЕМ МАШИНОВЕДЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК»
- ФГУП «ГОСУДАРСТВЕННЫЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ АВИАЦИОННЫХ СИСТЕМ»
- ФГБУН «ИНСТИТУТ ПРОБЛЕМ УПРАВЛЕНИЯ ИМ. В. А. ТРАПЕЗНИКОВА РАН»
- ГНЦ ФГУП «ЦЕНТРАЛЬНЫЙ АЭРОГИДРОДИНАМИЧЕСКИЙ ИНСТИТУТ ИМЕНИ ПРОФЕССОРА Н.Е. ЖУКОВСКОГО»
- ФГАОУ ВО «САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ПЕТРА ВЕЛИКОГО»
- ФГАОУ ВО «САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ»
- ФГБВОУ ВО «ВОЕННО-КОСМИЧЕСКАЯ АКАДЕМИЯ ИМЕНИ А.Ф. МОЖАЙСКОГО»
- ФГАОУ ВО «МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМ. Н.Э. БАУМАНА (НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»
- ВУНЦ ВМФ «ВОЕННО-МОРСКАЯ АКАДЕМИЯ ИМ. АДМИРАЛА ФЛОТА СОВЕТСКОГО СОЮЗА Н.Г. КУЗНЕЦОВА»
- ФГУП «КРЫЛОВСКИЙ ГОСУДАРСТВЕННЫЙ НАУЧНЫЙ ЦЕНТР»
- ФГБУН «ИНСТИТУТ ОКЕАНОЛОГИИ ИМ. П.П. ШИРШОВА РАН», САНКТ-ПЕТЕРБУРГСКИЙ ФИЛИАЛ
- ЗАО «МОРСКИЕ НЕАКУСТИЧЕСКИЕ КОМПЛЕКСЫ И СИСТЕМЫ»

ISBN 978-5-91995-059-2
© АО Концерн ЦНИИ Электроприбор, 2018

СОПРЕДСЕДАТЕЛИ ПРОГРАММНОГО КОМИТЕТА КОНФЕРЕНЦИИ:

ЧЛЕН-КОРРЕСПОНДЕНТ РАН Р.М. ЮСУПОВ,
ЧЛЕН-КОРРЕСПОНДЕНТ РАН В.Н. ВАСИЛЬЕВ,

ДОКТОР ТЕХНИЧЕСКИХ НАУК, ПРОФЕССОР В.Н. ШЕЛУДЬКО

ЗАМЕСТИТЕЛЬ ПРЕДСЕДАТЕЛЕЙ ПРОГРАММНОГО КОМИТЕТА КОНФЕРЕНЦИИ:

ДОКТОР ТЕХНИЧЕСКИХ НАУК, ПРОФЕССОР Б.В. СОКОЛОВ

ЧЛЕНЫ ПРОГРАММНОГО КОМИТЕТА КОНФЕРЕНЦИИ:

В. В. Александров
Т. И. Алиев
Д. Г. Арсеньев
А. А. Бобцов
С. Н. Васильев
Д. В. Гайворонский
В. И. Городецкий
С. Ю. Желтов
П. Д. Зегжда
Я. А. Ивакин
П. М. Иванов
Б. Г. Ильясов
И. А. Каляев
А. А. Карпов
О. В. Карсаев
В. В.Касаткин
В. Н. Козлов
И. В. Котенко
Г. А. Леонов
О. Б. Макаревич
А. А. Молдовян
Д. А. Новиков
В. Ю. Осипов
Ю. С. Попков
В. В. Путов
В. Х. Пшихопов
А.А. Ронжин
Г. Г. Себряков
П. А. Скобелев
А. В. Смирнов
Б. Я. Советов
В. А. Сойфер
Б. В. Соколов
А. Г. Степанов
О. А. Степанов
В. А. Терехов
Б. Ф. Фомин
А.Л. Фрадков
М. Ю. Шестопалов
В. П. Шкодырев

С О Д Е Р Ж А Н И Е

Секция 1
ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В УПРАВЛЕНИИ

Юсупов Р.М., Мусаев А.А.
Проблема оценивания эффективности информационных технологий 11

Козлов В. Н., Ефремов А. А.
Проекционные операторы математического программирования для интеллектуальной оптимизации управлений динамическими объектами 21

Косовская Т. М.
Построение многоуровневой базы для уменьшения вычислительной сложности решения задачи конъюнктивный булевский запрос 31

Осипов В. Ю., Жукова Н. А., Червонцев М. А., Водяхо А. И., Климов Н. В.
Когнитивные технологии в управлении мониторингом. 37

Алёшин Е. Н., Терещук В. А., Терещук С. А.
Модели интеллектуальной системы поддержки принятия решений по управлению движением космического аппарата при возникновении нештатных ситуаций.....46

Макшанов А. В., Мусаев А. А.
Анализ эволюционирующих спектров нестационарных процессов 54

Баженов А. Г., Шарыгин Б. Л.
Опыт формирования электронного определения изделия на всех этапах его жизненного цикла на примере двухстепенного поплавкового гироскопа 60

Афанасьева И. В., Левоневский Д. К., Новиков Ф. А., Федорченко Л. Н.
Формальная модель и методы описания систем со сложным поведением на принципах обобщённых автоматов 67

Харитонов Н. А.
Алгоритм построения неоднородной математической структуры алгебраических байесовских сетей 77

Секция 2
СИСТЕМНЫЙ АНАЛИЗ И КОМПЛЕКСНОЕ МОДЕЛИРОВАНИЕ АВТОМАТИЗИРОВАННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

Лукинский В. С., Искандеров Ю. М., Соколов Б. В., Некрасов А. Г.
Проблемы и перспективы использования интеллектуальных информационных технологий в логистических системах. 80

Дорогов А. Ю.
Моделирование протоколов маршрутизации Manet сетей 90

Алферов В. В., Кондратьев В. В. Соколов Б. В., Охтилев М. Ю., Салухов В. И.
Принципы и подходы к использованию сервисно-ориентированной архитектуры для формирования интеллектуального информационного пространства предприятия 101

Иванов В. П.
Метод адаптации оптимального управления для динамических систем 110

Алексеев А. В., Мусатенко Р. И., Тычинин И. Ю.
Системный мониторинг при интеллектуальной поддержке управления критических объектов 119

Кулаков А. Ю., Кулаков Ф. М., Павлов А. Н., Потрясаев С. А.,Соколов Б. В.
Модельно-алгоритмическое обеспечение планирования реконфигурации бортовой аппаратуры маломассоразмерных космических аппаратов 123

Качанова Т. Л., Фомин Б. Ф., Фомин О. Б., Агеев В. О.
Типология системных эффектов многофакторных воздействий в открытых системах 131

Шишкин В. М., Колесников К. Е.
Управление конфронтационным взаимодействием в условиях ресурсообмена сторон – динамическая модель 140

Потапов С.Е.
Методика обоснования параметров дисциплины обслуживания разнородных абонентов мультисервисной сети доступа при заданных требованиях к характеристикам информационного обмена 149

Секция 3
МУЛЬТИАГЕНТНЫЕ СИСТЕМЫ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И НЕЙРОТЕХНОЛОГИИ В ЗАДАЧАХ УПРАВЛЕНИЯ

Карсаев О. В., Мартынова Л. А.
Групповое управление мультиагентной системой автономных необитаемых подводных аппаратов при ведении всепогодной сейсморазведки 155

Осипов В. Ю., Никифоров В. В.
Потоковые рекуррентные нейронные сети с управляемыми элементами 166

Пшихопов В. Х., Медведев М. Ю., Пархоменко В. Б., Васильева М. А.
Исследование интеллектуальной системы управления автономного наземного робота с нейронной сетью глубокого обучения 176

Скобелев П. О., Симонова Е. В., Будаев Д. С., Вощук Г. Ю., Ларюхин В. Б.
Облачная интеллектуальная система SMART FARMING для управления точным земледелием 183

Колпаков В. Е.
Перспективы использования нейронных сетей для управления системой самодиагностики судовой энергетической установки 193

Блюм В. С.
Методы цифровизации информационного потока в клинической сфере здравоохранения 200

Власенко С. В.
Анализ эффективности применения аппарата концептуальных графов в задачах интеллектуальной обработки моделей цифровых производств 210

Арсеньев Д. Г., Баскаков Д. Е., Шкодырев В. П.
Контрактные сетевые протоколы взаимодействия когнитивных агентов в распределенных интеллектуальных системах управления 218

Секция 4
ПРИКЛАДНЫЕ ЗАДАЧИ ОБРАБОТКИ ИНФОРМАЦИИ В УПРАВЛЕНИИ

Акулов В. А.
Технология виброакустического мониторинга городских автобусов с элементами мультиагентности 223

Гриняк В. М., Девятисильный А. С., Цыбанов П. А.
Оценка точности навигации внутри помещений по данным Bluetooth устройств.....231

Харинов М. В.
Модернизация информационных технологий в области обработки изображений.....240

Жвалевский О. В.
Построение многоуровневой системы распознавания для диагностики болезни Паркинсона 247

Жвалевский О. В.
Применение методов препроцессинга временных рядов для обработки тензотреморограмм 256

Алексеев А. В., Бобрович В. Ю., Антипов В. В., Смольников А. В.
Информационная технология и система мониторинга конкурентной способности продукции как главного фактора технологического развития предприятия.....265

Ханыков И. Г.
Трехэтапная блок-схема сегментации изображений.....270

Яковис Л. М., Доронина Н. А.
Учет характера параметрической неопределенности динамической модели при расчете робастных регуляторов для объектов с самовыравниванием 280

Колесникова С. И., Цветницкая С.А., Побегайло П.А
Обобщение метода агрегированных регуляторов на стохастический дискретный объект и его применение для решения прикладных задач.....291

Имаев Д. Х., Немудрук М. Л., Федоров М. С., Шестопалов М. Ю.
Интеллектуальные системы удаленного диагностирования технического состояния агрегатов 300

Секция 5
ОБРАБОТКА СИГНАЛОВ В СИСТЕМАХ УПРАВЛЕНИЯ И СВЯЗИ
(СЕКЦИЯ ПАМЯТИ А.Е. БАРАБАНОВА)

Задорожний В. Г., Курина Г. А.
О периодических в среднем решениях линейных неоднородных дифференциальных уравнений со случайными коэффициентами.....310

Бондарко В. А.
Относительный порядок дискретных моделей линейных стационарных систем 320

Пацко В. С., Федотов А. А.
Множество достижимости для машины Дубинса 325

Салищев С. И.
Разработка и оценка эффективности малопотребляющих полупроводниковых схем для цифровой обработки сигналов на основе антимашин Хартенштейна 334

Викулов Е. И., Салищев С. И.
Адаптация функции корреляции каналов в микрофонных решетках для улучшения качества распознавания речи при помощи постфильтра шумоподавления 345

Богаевский Д. В., Ежов С. Н., Каплун Д. И., Кошкин А. Д., Аряшев С. И.
Исследование возможностей применения перспективных отечественных векторных процессоров при обработке сигналов и многомерной визуализации 347

Сомов Е. И., Бутырин С. А., Сомова Т. Е.
Алгоритмы калибровки и дискретной фильтрации сигналов в астроинерциальной системе для определения ориентации маневрирующего спутника землеобзора 358

Томашевич С. И., Зенкин А. М.
Кодирование к аудио- и видеосигналов с помощью адаптивного кодирования 368

Половцев С. Е., Клионский Д. М.
Разработка и программная реализация имитатора сигналов 374

Секция 6
БЕЗОПАСНОСТЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В УПРАВЛЕНИИ

Паращук И. Б., Агеев С. А.
Повышение достоверности оценки смыслового наполнения информационных объектов на основе обработки неполных, противоречивых и нечетких знаний 378

Котенко И. В., Паращук И. Б.
Общая архитектура интеллектуальной системы аналитической обработки цифрового сетевого контента в интересах защиты от нежелательной информации 384

Паращук И. Б.
Показатели качества системы обеспечения безопасности информационных технологий в управлении и оценивание эффективности ее функционирования 389

Кушнеревич А. Г., Браницкий А. А.
Программное средство распределенного выполнения процедур машинного обучения для целей сетевой безопасности 394

Виткова Л. А., Дойникова Е. В.
Поддержка принятия решений по противодействию нежелательной информации 398

Десницкий В. А., Браницкий А. А.
Принципы адаптации и переобучения системы анализа информационных объектов для защиты от нежелательной и вредоносной информации 404

Дойникова Е. В
Подход к прогнозированию и реагированию на кибератаки в индустриальном Интернете вещей на основе нейронечётких сетей 408

Дойникова Е. В., Федорченко А. В.
Применение онтологического хранилища для анализа защищенности информационных систем 413

Хакимова Э. Р., Чечулин А. А.
Методика защиты детей с ограниченными возможностями от нежелательной информации в сети Интернет 418

Гайфулина Д. А., Федорченко А. В.
Методика предварительной обработки сетевого трафика для обнаружения аномалий в кибер-физических системах 428

Меркушев Е. С.
Механизмы обеспечения целостности данных в облачных системах, основанные на протоколах доказательства обладания данными 433

Коломеец М. В., Левшун Д. С.
Визуализация данных в интересах выявления и противодействия нежелательной и сомнительной информации 437

Левшун Д.С., Бахтин Ю.Е.
Эмерджентные свойства киберфизических систем 441

Комашинский В. И.
Методы профилирования и выявления аномалий для обнаружения вредоносной активности 446

Саенко И. Б., Комашинский В. И.
Система показателей и критериев оценки качества политик разграничения доступа в облачных инфраструктурах критически важных информационных систем 451

Кулаков А. Д., Левшун Д. С.
Горячее подключение внешних электронных компонентов к устройствам платформы Arduino 457

Старков А. М., Саенко И. Б.
Подход к разработке методического аппарата для технологического управления виртуальными вычислительными сетями в корпоративных информационных системах 461

Бахтин Ю. Е., Левшун Д. С.
Реализация полунатурной модели железной дороги для анализа инцидентов безопасности 469

Агеев С. А., Саенко И. Б.
Оперативное управление маршрутизацией в мультисервисных сетях связи с учетом обеспечения информационной и сетевой безопасности при неполноте контролируемой информации 474

Котенко И. В., Саенко И. Б.
Концептуальные основы построения интеллектуальных систем аналитической обработки цифрового сетевого контента с целью обнаружения и противодействия нежелательной, сомнительной и вредоносной информации 482

Десницкий В. А., Рудавин Н. Н.
Подход к анализу защищенности беспроводной самоорганизующейся сети кризисного управления 488

Хлобыстова А. О., Абрамов М. В., Тулупьев А. Л.
Идентификация наиболее вероятных траекторий социоинженерных атак в управлении рисками, ассоциированными с пользователями 493

Шаламов Р.А., Абрамов М.В., Тулупьева Т.В.
Автоматизация оценки степени выраженности психологических особенностей пользователей в управлении кадрами: тестирование в социальной сети.....497

Секция 7
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ЧЕЛОВЕКО-МАШИННОГО ВЗАИМОДЕЙСТВИЯ

Гураков И. А., Костюченко Е. Ю., Новохрестова Д. И., Шелупанов А. А.
Статистические распределения формант различных дикторов при проведении фоноскопических экспертиз методом формантного выравнивания.....501

Анисимов Я. О., Кацай Д. А., Исупова Т. Д.
Алгоритмический подход к определению интонационной характеристики звукового образца 510

Марковников Н. М., Кипяткова И. С.
Исследование методов построения интегральных систем распознавания русской речи 518

Аксёнов А. А., Рюмин Д. А., Иванько Д. В.
Разработка системы геометрических признаков для автоматического распознавания визуальной русской речи.....526

Величко А. Н., Карпов А. А., Будков В. Ю.
Аналитический обзор речевых корпусов для систем определения ложных речевых сообщений 534

Маркитантов М.В., Карпов А. А.
Аналитический обзор подходов к автоматическому распознаванию возраста диктора по голосу 539

Ненаусников К. В.
Подходы к построению ассоциативных словарей для повышения точности анализа естественно-языковых текстов 543

Верхоляк О. В.
Аналитический обзор информационного обеспечения для автоматического определения тональности русскоязычных текстов 548

Котенко И. В., Коломеец М. В., Бушуев С. Н., Гельфанд А. М.
Методы человеко-машинного взаимодействия на основе сенсорных экранов в ситуационных центрах безопасности 554

Коломеец М. В., Котенко И. В., Косов Н. А., Агеев С. А., Иванов А. Ю.
Анализ методов человеко-машинного взаимодействия в инструментах визуальной аналитики SIEM-систем.....559

Кагиров И. А.
Аналитический обзор робототехнических средств с многомодальным человеко-машинным интерфейсом.....563

Крючков Б. И., Усов В. М., Иванько Д. В.
Перспективы использования интеллектуальных пространств для информационной поддержки оператора при дистанционном мониторинге группы мобильных роботов на лунной поверхности 572

Михайлюк М. В., Крючков Б. И., Усов В. М.
Разработка бортового виртуального симулятора для формирования и поддержания готовности экипажа на лунной станции к аварийному реагированию на потенциально опасные ситуации 582

Секция 8
ПРОБЛЕМЫ ИНЖЕНЕРНОГО ОБРАЗОВАНИЯ В ОБЛАСТИ УПРАВЛЕНИЯ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ: СОВРЕМЕННОЕ СОДЕРЖАНИЕ И ТЕХНОЛОГИИ ОБУЧЕНИЯ

Степанов А. Г.
Профессиональная подготовка или образование? 591

Боронахин А. М., Ткаченко А. Н., Шалымов Р. В.
Вопросы обеспечения качества образовательного процесса в современных условиях 595

Фаткиева Р. Р., Константинова К.С.
Применение проектной деятельности при обучении в ВУЗе 605

Котова Е. Е.
Вопросы подготовки IT-специалиста в условиях цифровой трансформации 610

Котова Е. Е., Писарев А. С., Писарев И. А.
Технологии автоматизированной обработки и интеллектуального анализа неструктурированных данных 619

Секция 1
ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В УПРАВЛЕНИИ

ПРОБЛЕМА ОЦЕНИВАНИЯ ЭФФЕКТИВНОСТИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Р. М. Юсупов[1], **А. А. Мусаев**[2]
(*СПИИРАН, Санкт-Петербург, 199178, РФ, С.-Петербург, 14 линия ВО, 39, spiiran@iias.spb.ru*)

Аннотация

Рассмотрены проблема и задача оценивания эффективности информационных систем и технологий, позволяющая количественно проанализировать результативность их применения. Предложены количественные оценки эффективности информационной технологии, учитывающие вероятностную природу исходных данных, используемых для построения этих оценок. Приведены варианты необходимых и достаточных критериев эффективности информационных технологий, выраженные через эффективность метасистем, в интересах которых они функционируют.

Ключевые слова: информация, информационные технологии, эффективность, методология, вероятностный подход, системный анализ.

Введение

Процесс создания и применения любой *информационной технологии* (ИТ) неизбежно связан с проблемой оценки их эффективности. При этом возникает целый ряд сложных вопросов, обусловленных спецификой объекта, для переработки и хранения которого ИТ создавалась – *информации*. Над этим понятием многие годы трудились философы и прикладные специалисты, однако их успехи существенно отстают от взрывоподобного роста практических приложений ИТ. В результате возникшего разрыва между теоретическим осознанием природы информационных процессов и быстро прогрессирующим развитием практических реализаций ИТ образовался пробел в области оценки их эффективности, связанный с общей в задачей квалиметрии информационных систем.

Особую проблему при оценивании эффективности ИТ создает тот факт, что большинство таких технологий относится к сервисным или инструментальным средствам. Это означает, что сами по себе они не создают итогового результата, который можно было бы оценить в терминах экономической эффективности. Иными словами, результативность ИТ является опосредованной через функционирование других видов систем - технических, экономических, военных, социальных и т.п.В связи с этим возникает задача формирования методологической базы, увязывающей оценку эффективности ИС с общей теорией системной квалиметрии. Решению данной задачи и посвящена настоящая работа.

Состояние проблемы и оценивания эффективности информационных систем

Отсутствие единого подхода к задаче оценивания эффективности ИТ является закономерным следствием диалектического противоречия между общим и частным. Общий подход слабо отражает существенные особенности информационных систем, частные технологии оценки эффективности отдельных решений не подходят для других приложений.

Информационные технологии и системы обычно также относятся к инновационным решениям и, теоретически, они должны быть оценены на основе существующих методик. Однако defacto применение экономических методик анализа инноваций к связано с существенными трудностями. Главной причиной этого является нематериальный характер объекта приложений – информации. В частности, следствием нематериального характера информационного «сырья»

[1] Доктор технических наук, член-корреспондент РАН, научный руководитель СПИИРАН.
[2] Доктор технических наук, главный научный сотрудник.

является косвенный характер влияния информационных инноваций на экономическую эффективность проектов.

Проблемы оценивания эффективности начинаются с неоднозначности исходных определений. В различных отечественных и зарубежных руководящих документах используются разнообразные определения эффективности [1–5]. По определению ISO 9000, «эффективность — это связь между достигнутым результатом и использованными ресурсами» [2]. При этом, как замечено в [6], перевод с английского, осуществленный в процессе преобразования ISO 9000 в ГОСТ Р ИСО 9000, содержит некорректную интерпретацию терминов *efficiency* и *effectiveness*. В частности, не учитываются содержательные различия в терминах «результативность» и «производительность».

Часто используется еще одно плохо формализуемое определение – «эффективность инвестиционного проекта — категория, отражающая соответствие инвестиционного проекта целям и интересам его участников» [1]. Относительно конструктивное определение эффективности, привязанное к экономическому анализу, можно найти в [5]: «Под экономической эффективностью понимается соотношение непосредственных результатов деятельности, результатов, полученных и планируемых для достижения в рамках тактических задач, программ, программных мероприятий, с затратами на их достижение».

Возникает закономерный вопрос – можно ли на основе таких размытых определений построить однозначную количественную оценку эффективности (или любой другой системы)? Вопрос открытый. Сложность решения данной задачи привела к скептическому отношению относительно возможности априорных оценок эффективности инновационных проектов в целом.

В частности, авторитетный американский журнал CIOInsight утверждает, что более половины фирм вообще не осуществляют анализ окупаемости вложений в развитие информационной структуры (даже при внедрении столь существенных проектов, как ERP или корпоративные порталы). Вердикт журнала: «Состояние оценивания экономического эффекта от внедрения ИТ выглядит удручающе» [8]. По результатам опроса российских предпринимателей [9], в целом, сохраняется вера в окупаемость вложений в информационную инфраструктуру компании (до 47% опрошенных). Однако 80% компаний, из числа не согласившихся на внедрение ИТ, в качестве причины отказа называют недостаточность эффекта от вложений.

Насколько эффективны современные западные методики оценки эффективности инвестиционных проектов? Речь идет, прежде всего, о таких интегральных показателях, как *TCO* (totalcostofownership, полная стоимость владения)и *ROI* (returnoninvestment, возврат на инвестицию). По мнению ряда авторов [8,9], отказ российских предпринимателей от использования этих показателей связан с иной ориентацией отечественного бизнеса. В РФ основной целевой функцией является не увеличение производительности труда, а рост отдачи от капитала.

Для ИТ-инноваций оба приведенных показателя являются несостоятельными в силу сложности оценки парциальной прибыли, обусловленной внедрением или модификацией информационных систем [10]. К тому же высочайшая скорость развития вычислительных средств крайне затрудняет оценку срока использования . По результатам опроса [11], 70% компаний находят показатель ROI слишком сложным для вычисления и 44% – субъективным.

Следуя [4, 6], рассмотрим еще одну модель – *PRM*, применяемую, в частности, для оценки эффективности ИТ при использовании в интересах государственных служб. В качестве основного показателя Productivity&Efficiency используется объем работы, выполненный в единицу времени и соотнесенный с использованными ресурсами.

Важные результаты в отношении возможности применения традиционных методик оценки эффективности инноваций к ИТ показали исследования П. Страссмана [12, 13]. Вывод оказался достаточно неожиданным: между ИТ-бюджетами и целевыми показателями результативности бизнеса статистических корреляций нет! Исключение составил показатель затрат на управление и администрирование, продвижение и поддержку продаж – Sales, General&Administrative (*SG&A*). Заметим, что данный вывод подтверждает справедливость вышеприведенного утверждения о косвенном характере влияния ИТ на терминальную эффективность бизнеса. На основе приведенных исследований в качестве оценки реальной эффективности был предложен показатель информационной продуктивности – *InformationProductivity* (IP) [12–15] *IP = EVA/ SG&A,* где *EVA* (EconomicValueAdd) – объем прибыли компании после уплаты налогов и вы-

платы дивидендов по акциям. Здесь *SG&A* используется как параметр, чувствительный к внедрению ИТ. Анализ динамики *IP* показал, что лишь немногие компании смогли получить конкурентные преимущества за счет внедрения ИТ. Анализ этого явления приведен в [14].

В качестве альтернативы к бизнес-оценкам эффективности в [10] приводятся так называемые системные методы анализа, демонстрирующие целесообразность применения ИТ в различных сферах деятельности компаний. Примерами соответствующих методик могут служить:

– сбалансированная система показателей *BSC* (BalancedScoreCard);
– системапоказателейИТ*BITS* (Balanced IT Score Card);
– референтная модель производительности *PRM* (PerformanceReferenceModel);
– бизнес-ценностьИТ*BVIT* (Business Value of IT);
– сервиснаястратегия*ITIL* (Information Technology Infrastructure Library Service Strategy).

Существенным отличием системных методик является попытка отразить не только терминальные финансовые показатели, но и другие характеристики, благоприятно влияющие на деятельность компании. Примерами таких характеристик являются:

– вклад ИТ в развитие компании;
– операционные преимущества;
– влияние на перспективы компании;
– степень ориентации на клиентов и пользователей;
– улучшение качества предоставляемых услуг;
– степень влияния на бизнес-процессы;
– прямая окупаемость ИТ-инноваций и т.п.

Очевидными недостатками всех приведенных определений является их низкая пригодность для построения точных количественных оценок эффективности на разных этапах их жизненного цикла. Достаточно понятно, что построить гармоническую интегральную оценку эффективности ИТ-инноваций на таком обширном, плохо формализованном пространстве практически невозможно. Результатом применения размытых методик, в конечном итоге, является множество числовых характеристик, не допускающих ранжирования, а, следовательно, и корректного сравнения вариантов реализации без вмешательства субъективной экспертизы.

Следует заметить, что отсутствие объективных подходов к оценке эффективности ИТ, привело некоторых исследователей к парадоксальным выводам. Так, например, в работе Н. Карра«Блеск и нищета информационных технологий» [15] обосновывается вывод о том, что ИТ вообще не дают конкурентного преимущества. При этом приводятся вполне убедительные доводы: комплексная автоматизация предприятия на уровне ERP может стоить миллионы USD и потребовать существенной перестройки всей информационной инфраструктуры предприятия. В то же время сомнительный положительный эффект достигается только косвенным образом, через увеличение уровня транспарентности, и, как ожидаемого следствия, управляемости производством. При этом, в силу фантастической скорости развития микроэлектроники и ИТ, развернутая дорогостоящая ERP может оказаться устаревшей уже к моменту ее внедрения.

В работах [17–18] также указывается, что сами по себе ИТне дают преимуществ и лишь в совокупности с другими инициативами способны влиять на ключевые факторы успеха, создавать конкурентные преимущества. Для оценки эффективности в этих статьях предлагается использовать процессный подход, разделяющий корпоративные процессы на основной, т.е. добавляющий стоимость к исходному продукту, и вспомогательный. Для промышленных предприятий информационные процессы естественно относятся к категории вспомогательных. Следовательно, эффективность в этом случае также следует оценивать косвенно, через совершенствование управления производственного цикла предприятия.

Экспериментальное опровержение парадоксу непроизводительности ИТ приведено в работах группы экспертов под руководством Э. Бринйолфсона [19–20]. В частности, было показано, что ИТ оказывают существенное и статистически значимое влияние на итоги работы компании, отдача на инвестиции в компьютерный капитал составляла в среднем около 81% и существенно превосходила отдачу от инвестиций в другие материальные активы.

В заключении обзора, посвященного эффективности, следует указать на материалы проекта IT-Value [21], описывающие особенностям экономического анализа инвестиций в ИТ. Исходными постулатами проекта являются утверждения:

– инвестиции в ИТ влияют на всю организацию, поэтому объект исследования – это компании в целом, а не информационные системы или ИТ-проекты по отдельности;

– развитие и эксплуатация корпоративных ИТ рассматривается как единый и взаимосвязанный комплекс процессов, который влияет на бизнес-результаты компании;

– существуют объективные закономерности влияния ИТ на экономические результаты компании, которые не зависят от того, что думают о них руководители организации или ИТ-директора и др.

Важной особенностью проекта является разработка методов оценки эффективности инвестиций в ИТ, учитывающих специфику российской действительности.

Кратким выводом из приведенного обзора является заключение об актуальности задачи построения конструктивной количественной методики анализа эффективности , обеспечивающей априорную оценку их результативности и возможность объективного сравнения вариантов их реализации с учетом нематериальной специфики информационных сред. В качестве теоретической платформы для построения такой методики предлагается проверенное сочетание вероятностно статистического подхода к оцениванию эффективности с общими принципами системного анализа [22–27].

Базовые понятия теории эффективности информационных систем и технологий

Выбор критериальной функции, предназначенной для количественной или качественной оценок результативности , достаточно субъективен. Наиболее универсальным показателем качества системы является *эффективность* ее функционирования.

Эффективность является одним из фундаментальных свойств любой системы и непосредственно сопоставляется с результативностью ее применения. В связи с этим понятие эффективности системы, во многих случаях, используется в качестве основного показателя качества ее функционирования.

Для описания столь сложного и многопланового понятия, как эффективность, применяется целый набор разнородных показателей, отражающих определенные квалиметрические признаки. В частности, показатели эффективности различают по степени их интегральности (скалярные, векторные), по типу шкал (количественные, качественные), по степени достаточности (необходимые, достаточные), по характеру наблюдений (прямые, косвенные) и т.п.

Принципиальным вопросом для оценки эффективности является размерность ее показателя: одномерность (или скалярность) интегрального показателя эффективности является существенной. В этом случае сравнительный анализ вариантов решения поставленной задачи осуществляется их прямым ранжированием на основе очевидного упорядочивания значений скалярного интегрального показателя эффективности.

В случае отказа от одномерного показателя эффективности и перехода к векторному, элементами которого являются частные характеристики формируемого решения, задача выбора наилучшего варианта не только существенно усложняется, но и становится не вполне корректной. Действительно, сравнение и упорядочивание ряда векторов, отвечающих набору возможных решений, реализуется лишь путем их скаляризации на основе той или иной метрики. В свою очередь, выбор подходящей метрики всегда несет в себе элементы субъективизма. В лучшем случае можно ожидать выбор метрики на основе экспертных оценок, когда весовые коэффициенты частных показателей определяются специальной обработкой результатов опроса группы или групп специалистов в соответствующих областях знаний.

Тем не менее, векторные показатели эффективности также распространены на практике, поскольку позволяют осуществить естественную декомпозицию общей проблемы повышения эффективности функционирования на ряд частных вопросов, каждый из которых может составлять самостоятельную научную задачу последующего иерархического уровня. Так, например, комплексная проблема совершенствования состоит из задач повышения оперативности доступа к хранимой информации, уровня программной совместимости с другими *системами хранения информации* (СХИ) и с информационными приложениями, степени безопасности хранения данных, степени доступности данных для пользователей, возможности формирования нестандартных запросов и т.д. Необходимость учета частных показателей приводит к специфическим методам анализа и синтеза систем, основанным на многокритериальном подходе.

На вербальном уровне эффективность определяется как количественная мера способности системы выполнять возложенные на нее задачи. Прямой эффект от применения конкретной системы оценивается так называемой *функциональной эффективностью*. Наряду с функциональной эффективностью широко используется понятие *экономической эффективности*, в котором достигнутый полезный эффект соотносится с экономическими затратами, потребляемым ресурсом и т.п. Окончательный выбор показателя и критерия эффективности осуществляется, исходя из цели, в интересах которой создавалась эта система. Так, например, АСУ, рассматриваемая как информационная система, создавалась с целью повышения эффективности формируемых управляющих решений и их реализации, образовательной системы – с целью повышения качества подготовки выпускников и т.п.

Таким образом, основные показатели эффективности формируются на основе требований, предъявляемых к ней со стороны *метасистемы,* под которой понимается иерархически вышестоящая система, в интересах которой создается.

Количественная оценка эффективности информационной системы.

Основные технические характеристике любой ИС $IS(Q)$ задаются векторной совокупностью ее собственных показателей $Q=(q_1, q_2,..., q_n)$.

Для ИС в роли таких показателей могут выступать:

– допустимый объем хранения полезной информации;

– реальный объем памяти hardware, занимаемый ИТ;

– время доступа (время получения пользователем единицы объема востребованных данных из СХИ);

– степень безопасности ИТ в отношении несанкционированного доступа;

– степень надежности сохранения информации в СХИ;

– открытость системы и т.п.

Как видно из приведенного перечня, собственные показатели эффективности ИТ могут обладать различной природой: часть из них носит количественный (метрический) характер, хорошо согласующийся с традиционными методами статистического оценивания, другие имеют качественный характер и выражаются через логические переменные. В последнем случае требуются специальные методы обработки, основанные на правилах символьных исчислений, неметрическом шкалировании, теории нечетких множеств и т.п. В результате возникает проблема согласования измерительных шкал (метрической, ординальной, номинальной), решение которой позволяет использовать разнородные исходные данные и довести решение задачи анализа эффективности ИТ до конструктивного результата.

Как правило, технические требования к ИТ задаются в виде области допустимых значений $\{Q_{IS}\}_0$, полученной в результате априорного анализа аналогов, экспериментального и/или имитационного исследования функционирования информационной системы IS и метасистемы MS, в интересах которой она была создана. Наличие априорных требований к ИС позволяет сформировать простейший *критерий пригодности ИС* в виде:

$$Q_{IS} \in \{Q_{IS}\}_0 . \qquad (1)$$

Как отмечалось, на практике в критерий (1) подставляется не значения вектора Q, а его оценка $\hat{Q}$, полученная по результатам измерений или априорного экспертного анализа. Следовательно, критерий (1) носит статистический характер.

Более общий подход основан на предположении о том, что множество допустимых значений технических характеристик ИС $\{Q_{IS}\}_0$ должно отвечать требованию пригодности иерархически вышестоящей метасистемы MS, в интересах которой информационная система создавалась. Предположим, что эффективность (результативность, достигнутый положительный эффект) метасистемы MS определяется скалярной величиной Eff_{MS}, зависящей от технических характеристик входящей в ее состав ИТ с параметрами Q_{IS}, т.е.

$$Eff_{MS} = Eff_{MS}(Q_{IS}) . \qquad (2)$$

Реальная эффективность функционирования метасистемы MS зависит от множества разнообразных факторов: показателей самой метасистемы, качества ее управления, особенностей применения и т.п. Однако для решения задачи оценки эффективности ИТ интерес представляет только частная зависимость (2); остальные зависимости и параметры рассматриваются как условия, фиксируемые в ограничительной или дескриптивной части решаемой задачи.

Как уже отмечалось, оценка эффективности ИТ осуществляется исходя из результативности функционирования (применения) метасистемы. Оценка результативности метасистемы, рассматриваемая здесь как синоним ее эффективности, в свою очередь, определяется на основе субъективных критериев, имеющих экзогенный (по отношению к MS) характер. Следовательно, интегральная оценка эффективности ИТ Eff_{IS} определяется набором a'priori заданных критериев, позволяющих определить необходимые и достаточные условия эффективности метасистемы MS в зависимости от собственных показателей ИТ.

Необходимым условием эффективности любой системы является ее *пригодность.* Однако, в отличие от (1), как это только что было декларировано, она должна выражаться через терминальную эффективность метасистемы MS.

При использовании детерминированного подхода, когда в наличии имеется одна единственная реализация (или в условиях полной статистической определенности) критерий пригодности описывается соотношением

$$Eff_{MS}[Q_{IS}] > Eff_0 \ , \tag{3}$$

Здесь Eff_0 – некоторое пороговое значение эффективности метасистемы, ниже которого MS теряет функциональную пригодность. На практике оценка эффективности $Eff_{MS}[Q_{IS}]$ является случайной величиной. Следовательно, для повышения достоверности критерия пригодности (3) целесообразно перейти к ее усредненному значению, полученному на основе серии из N испытаний с одинаковыми условиями их проведения:

$$\overline{Eff_{MS}(Q_{IS})} = \frac{1}{N}\sum_{i=1}^{N} (Eff_{MS}(Q))_i > Eff_0 \ . \tag{4}$$

Совокупность $Q = Q_{IS}$, отвечающих условию (4), образует множество допустимых значений технических характеристик ИТ $\{Q_{IS}\}_0$.

В более общем случае, пригодность ИТ определяется вероятностью того, что достигнутый метасистемой положительный эффект (результат) будет превышать некоторый a'priori установленный порог $Eff^0_{MS} = Eff_0$. В этом случае критерий пригодности ИТ можно представить в виде:

$$P\{Eff_{MS}[Q_{IS}] > Eff_0\} > P_0 \, , \tag{5}$$

где P_0 – критическое значение, выбираемое исходя из требований решаемой метасистемой MS задачи.

Критерий пригодности является необходимым условием: его невыполнение делает бессмысленным разработку или применение соответствующей ИТ. Таким образом, выполнение критерия пригодности является некоторым императивом: с его помощью можно однозначно указать на нецелесообразность разработки/применения ИТ, но нельзя ответить на другой не менее важный вопрос – насколько эффективна ее реализация.

В связи с этим, как отмечалось в предыдущем разделе, в практике квалиметрии одновременно с (3–5) используются условия достаточности – критерии превосходства или оптимальности.

Критерий превосходства позволяет оценить ИТ на основе сопоставления ее параметров с параметрами системы-прототипа IS^0:

$$Eff_{MS}[Q(IS)] > Eff^0_{MS} \quad \text{для } \forall Q_{IS} \in \{Q_{IS}\}_0 \, , \tag{6}$$

где множество допустимых значений параметров ИТ $\{Q_{IS}\}_0$ формируется из условия выполнения требования пригодности MS (1). Величина $dEff_{MS} = Eff(Q) - Eff(Q^0)$ будет отражать выигрыш, достигнутый от применения ИТ IS вместо системы-прототипа IS^0.

Применение критерия (6) требует существенных оговорок:

1. Желательно, чтобы эффективность метасистемы является монотонно неубывающей функцией от параметров качества ИТ Q, то есть,

$$\text{если } Q_1 \geq Q_2, \text{ то } Eff_{MS}(Q_1) \geq Eff_{MS}(Q_2). \quad (7)$$

На практике, выполнение этого условия ожидать не следует. Повышение качества ИТ не гарантирует повышение эффективности MS.

2. Понятия "больше" или "меньше" при сравнении векторов определяются через построенные на них скалярные меры. Выбор меры представляет субъективный экзогенный процесс и требует дополнительного обоснования;

Очевидно, что критерий (6) не гарантирует, что исследуемая ИТ будет обладать наибольшей эффективностью. Поэтому более строгое решение может быть получено на основе *критерия оптимальности,* выраженного через эффективность метасистемы:

$$Eff_{MS}(Q^*) > Eff_{MS}(Q_{IS}) \quad \text{для } \forall Q_{IS} \in \{Q_{IS}\}_0. \quad (8)$$

При выполнении вышеприведенных ограничений 1–3, в частности, при выполнении условия монотонности (7), векторный критерий оптимальности можно представить в виде $Q = Q^* = extr\{Q_{IS}\}_0$, т.е. имеет наибольшее (или наименьше) значение среди $\forall Q_{IS} \in \{Q_{IS}\}_0$.

В ряде прикладных задач, связанных с оценкой эффективности ИТ, крайне плодотворной оказывается представление качественного состояния системы в виде точки *m*-мерного фазового пространства. При этом изменение эффективности ИТ будет соответствовать перемещению фазовой точки в данном пространстве $\{Q^m\}$. Попадание вектора Q в область $\{Q_0\}$ соответствует выполнению критерия пригодности, превышение значений вектора $q_j^0 \in \{Q_0\}, \quad j = 1,...,n$ – критерию превосходства и нахождение глобального максимума $q_j^*, \quad j = 1,...,n$ – критерию оптимальности.

Заметим, что приведенные выше показатели эффективности относятся к детерминированной ситуации. В случае, когда имеется возможность получить избыточный объем наблюдений показателей качества, целесообразно перейти к вероятностно-статистической постановке, обеспечивающей более реалистическую оценку эффективности ИТ. В частности, при наличии вероятностного распределения вектора Q, критерии пригодности, превосходства и оптимальности можно представить в виде, соответственно:

$$P\{Q \in \{Q_0\}\} > P_0; \quad (9)$$

$$P\{Q > Q^0\} > P_0; \quad (10)$$

$$P\{Q = Q^*\} > P_0. \quad (11)$$

При этом соотношение (9) является необходимым условием – ограничением для (10) и (11).

В более общем случае вероятностный подход требует знания функции распределения показателя эффективности метасистемы $F[Eff_{MS}(0)]$. По аналогии с необходимым критерием (критерием пригодности) (5) можно сформулировать и достаточные критерии эффективности. В частности, эффективность ИТ *по критерию превосходства* определяется вероятностью того, что достигнутый метасистемой положительный эффект $Eff_{MS}(Q(IS))$ будет превышать соответствующий эффект (полученный в тех же условиях), достигнутый при использовании ИТ-прототипа $IS^0 = IS(Q^0)$ с вероятностью, не ниже заданной P_0. В этом случае критерий превосходства ИТ можно представить в виде:

$$P\{Eff_{MS}[Q_{IS}] > Eff_{MS}[Q^0]\} > P_0, \quad (12)$$

При отсутствии полных априорных знаний о распределении параметра эффективности $F\{Eff_{MS}(Q)\}$, критерий превосходства можно сформировать на основе средних значений

$$\overline{Eff_{MS}(Q_{IS})} > Eff_{MS}(Q_{IS_0}) \text{ , для } \forall Q_{IS} \in \{Q_{IS}\}_0 \ . \quad (13)$$

Аналогичные соотношения могут быть получены для *критерия оптимальности*. При этом необходимо заранее определить множество, на котором осуществляется оптимизация. В частности, если осуществляется синтез или поиск оптимальной ИТ на некотором множестве допустимых значений показателей качества ИТ $\{Q_{IS}\}_0$, критерии оптимальности, в зависимости от уровня априорной статистической определенности, могут быть, по аналогии с (12 – 13), записаны в виде: $P\{Eff_{MS}[Q^*_{IS}] > Eff_{MS}[Q_{IS}]\} > P_0$, для $\forall Q_{IS} \in \{Q_{IS}\}_0$, где $Q^*(IS) = Q^*_{IS}$ – оптимальные значения параметров ИС.

Рассмотрим пример приложения предложенной методики к оценке эффективности автоматизированной системы управления (АСУ).

Пример. Оценка эффективности автоматизированной системы управления как информационной системы

Как уже отмечалось, выбор показателя и критерия эффективности ИТ осуществляется исходя из цели, определяемой со стороны иерархически вышестоящей метасистемы MS. В случае промышленного производства роль метасистемы MS, по отношению к АСУ, выполняет *технологический процесс* (ТП) или технологический цикл предприятия. Таким образом, эффективность АСУ, как информационной системы, следует оценивать через эффективность ТП.

Рассмотрим формализованную постановку оценки эффективности промышленной АСУ. Предположим, что ТП (т.е. метасистема *MS*), управляемый АСУ, функционирует в сложной, не полностью определенной среде под воздействием последовательностей управляющих воздействий $\{u_k, \ k = 1,...,m\}$. Состояние объекта управления $X_k, k = 1,...,N$, в общем случае, представляет собой векторный динамический стохастический процесс. Множество состояний, как правило, является континуумом, однако, в процессе взаимодействия с цифровой системой мониторинга оно преобразуется в дискретную временную последовательность.

Векторная совокупность текущего состояния ТП x_k, управляющих u_k и возмущающих v_k воздействий образуют текущую ситуацию $S_k = \{x_k, u_k, v_k, \ k = 1,...,N\}$.

Предположим, что $U = \{u_k\}$ – множество допустимых векторных управлений, формируемых АСУ. Каждое $u_k \in U$ представляет собой текущее управление, сформированное на основе анализа текущей ситуации S_k. В результате мониторинга состояния ТП образуется последовательность векторов наблюдения $Y_k = (y_1,...,y_{nv})_k, \ k = 1,...,N$, используемая для подготовки управляющих решений. Реализация управления приводит к реакции ТП, которая будет проявляться:

– в переходе ТП из состояния x_k в состояние x_{k+1} под воздействием управляющего сигнала $u_k \in U$;

– в выработке выходных воздействий (выходной продукции) $W_{k+\tau}$, определяющей терминальное качество управления.

В конечном итоге, ТП под воздействием управлений $U_N = \{u_1,...,u_N\}$ переходит из состояния x_1 в состояние x_N, в результате чего формируется положительный эффект $Э_N$ (функциональный или экономический, т.е. соотнесенный с затратами). Для одного производственного цикла эффективность управления будет оцениваться терминальным эффектом самого ТП, т.е. $Eff_{\text{IS}} = Э_N$. В случае, когда имеются данные по m завершенным производственным циклам, эффективность может быть уточнена традиционным усреднением $Eff_{\text{IS}} = \frac{1}{m}\sum_{i=1}^{m} Э_{\text{Ni}}$.

Рассмотрим задачу оценки эффективности пошагового корректирующего управления ТП. Пусть в результате теоретического анализа или моделирования установлено, что наибольшая достижимая величина положительного эффекта равна $Э^*_N$. Для достижения такого результата $Э^*_N$ известными методами оптимального управления формируется оптимальный план (программа) – последовательность управлений $U^*_N = \{u_1,...,u_N\}^*$. При этом, в соответствии с прин-

ципами системного анализа [22–24], эффект, достигаемый на каждом конкретном шаге может не быть наибольшим: важно, чтобы максимум достигался в конце управления при $k = N$ т.е. в момент достижения терминальной цели.

Поскольку объект управления находится под воздействием большого числа не полностью определенных возмущений, его состояние отклоняется от оптимальной фазовой траектории, обеспечивающей оптимальное решение поставленной задачи. Для демпфирования возникающих отклонений ИТ на основе обработки совокупности текущих наблюдений формирует последовательность корректирующих управлений $U_{\text{k}} = \{u_1,..., u_N\}_{k,} \quad k = 1,..., N$, позволяющую поддерживать реализацию оптимального плана.

В случае, если формируемая последовательность управления обеспечивает пригодное или оптимальное функционирование ТП, можно говорить, соответственно, о пригодности или оптимальности оцениваемой ИТ. Наоборот, если низкое качество управления приводит к тому, что система приходит в состояние $x_N \neq x_N^*$, то это означает неоптимальную работу ИТ, приводящую к дефициту результирующего эффекта $d\Э_N = \Э_N^* - \Э_N$. При этом, если $d\Э_N$ настолько велика, что полученный эффект оказывается меньше допустимого порога $\Э_0$, то ИТ оказывается не пригодной для выполнения поставленной перед ней задачей.

Отсюда, в соответствии с вышеприведенными критериями системной эффективности, можно получить конкретные формы оценки качества АСУ, как информационной системы.

ИТ является *эффективной по критерию пригодности*, если качество формируемой ей программы управления $Uo_N = \{u_1,...,u_N\}_0$ обеспечивает результирующий положительный эффект $\Э_N$ не меньше минимально допустимого $\Э_0$. Область эффективных (здесь – допустимых) программ управлений $\{U_0\} = \{u: \quad \Э_N \geq \Э_0\}$.

В случае, если имеется возможность набрать статистику управления, т.е. повторить процесс несколько раз, можно использовать вероятностные оценки эффективности. В этом случае АСУ, как ИТ, признается эффективной по критерию пригодности, если:

– результирующий положительный эффект $\Э_N$ в среднем оказывается не меньше минимально допустимого $\Э_0$, и, следовательно, область эффективных программ управлений $\{U_0\} = \{u: \quad \overline{\Э}_N \geq \Э_0\}$;

– вероятность превышения результирующим положительным эффектом $\Э_N$ минимально допустимого порога $\Э_0$ превышает заданную величину P_0. Соответственно, область эффективных программ управлений равна $\{\Э_0\} = \{\Э: \quad \text{P}\{\Э_\text{N} \geq \Э_0\} \geq \alpha_0\}$.

Аналогично описываются достаточные оценки эффективности *по критерию превосходства и по критерию оптимальности.*

Заключение

Основным выводом из предложенных материалов следует считать принципиальную возможность и необходимость количественного оценивания эффективности информационных систем и технологий через положительный эффект, достигаемый метасистемой, в интересах которой они созданы и применяются.

Следует заметить о наличии тесной взаимосвязью предложенного подхода с теорией риска. В частности, в качестве оценки эффективности ИТ по критерию пригодности предложено использовать вероятность превышения достигнутым положительным эффектом минимального порогового значения $P\{Eff_{MS}[Q_{IS}] > Eff_0\} \geq P_0$. Следовательно, последнее соотношение можно трактовать как вероятность выполнения стоящей перед ИТ задачи. Соответственно, дополнительная величина $Q_0 = 1 - P_0$ представляет собой вероятность невыполнения поставленной задачи, что совпадает с вероятностным определением понятия риска.

Работа проводилась при поддержке гранта РФФИ 12–07–00302–а.

ЛИТЕРАТУРА

1. Методические рекомендации по оценке эффективности инвестиционных проектов (Вторая редакция) // Минэкономики РФ, Минфин РФ, Госстрой РФ, 12.02.2016 // М.: Экономика. 2016. 421с.
2. ISO:9000:2015 fourth edition. Quality management systems. Fundamentals and vocabulary. – URL: https://www.iso.org/obp/ui/#iso:std:iso:9000:ed–4:v1:en
3. Международный стандарт МЭК 50 (191)–90. Надежность и качество услуг. Термины и определения. – URL: http://dic.academic.ru/dic.nsf/gosts/21964 (дата обращения: 26.11.2016).
4. FEA Consolidated Reference Model Document. Version 2.3. 2015. URL: http://www.slideshare.net/WillyDevNET/fea-crm-v23finaloct. (дата обращения: 26.11.2016).
5. Методические рекомендации по подготовке докладов о результатах и основных направлениях деятельности субъектов бюджетного планирования. Минфин РФ, МЭРТ РФ, июнь 2008. URL: https://www.google.ru/url?sa=t&rct=j&q=&esrc=s&source =web; (дата обращения: 26.11.2016).
6. Зиндер Е. Что такое «эффективность ИТ»? // IntelligentEnterprise. 2006. №8. – URL: https://www.iemag.ru/master-class/detail.php?ID=15727 (дата обращения: 26.11.2016).
7. CIO Insight. The voice of CIO community.URL:www.cioinsight.com. (дата обращения: 21.11.2016).
8. Арутюнян М. и др. Демистификация ИТ: Что на самом деле информационные технологии дают бизнесу / М. Арутюнян, Н. Ермошкин, С. Карминский и др. // М.: Альпина Бизнес Букс. 2006. 296 с.
9. Ермошкин Н.Н. Шесть способов оценить пользу и эффективность информационных технологий. // Банковский ритейл. 2006. №4. – URL: https://www.lawmix. ru/bux/76476(датаобращения: 21.11.2016).
10. Аншина М. Оценка эффективности ИТ. // Журнал школы IT-менеджмента «Системы управления бизнес-процессами». Академия народного хозяйства при Правительстве РФ. 2011. Выпуск № 7. – URL: http://journal.itmane.ru/node/591 (дата обращения: 20.11.2016).
11. ROI 2003: Do You Have Any Faith In Your ROI Numbers? – URL: http://www. cioinsight.com/c/a/Research/ROI–2003–Do–You–Have–Any–Faith–In–Your–ROI–Numbers#sthash. uvUbGa45.dpuf (датаобращения: 20.11. 2016)
12. Strassmann P.A. The squandered computer. Evaluating the business alignment of information technologies // New Canaan: Information Economics Press. 1997. 232p.
13. Pisello T., Strassmann P. IT value chain management. Maximizing the ROI from IT investment. // New Canaan: Information Economics Press. 2003. 236p.
14. Ананьин В.И. В поисках эффективности ИТ // IntelligentEnterprise. 2009. № 7,8,10. – URL: https://www.iemag.ru/analitics/detail.php?ID=18833 (дата обращения: 15.11. 2016).
15. Скрипкин К.Г. Экономическая эффективность информационных систем // М.: ДМК Пресс. 2003. 256с.
16. Карр Н. Блеск и нищета информационных технологий. Почему ИТ не являются конкурентным преимуществом //М: Секрет фирмы. 2005. 176с.
17. Новак Е.В. Информационно-коммуникационные технологии: оценка эффективности // Информационные технологии. 2014. № 8. С. 74–80.
18. Новак Е.В. Как измерить эффективность информационно-коммуникационных технологий // Статистика и экономика. 2014. № 3. С. 186–191.
19. Bresnahan T., Brynjolfsson E., Hitt L. Information Technology, Workplace Organisation and Demand for Skilled Labor: An Empirical Evidence // Quarterly Journal of Economics, 2002. Vol. 117. №1. P. 339–376.
20. Brynjofsson E., Hitt L., Yang S. Intangible Assets: Computers and Organization Capital // Brookings Papers on Economic Activity. 2002.№1. P. 137–198.
21. Эффективность инвестиций в ИТ. Альманах лучших работ // InformationManagement. 2012. 196с.
22. Андрейчиков А.В., Андрейчикова О.Н. Системный анализ и синтез стратегических решений в инноватике: Математические, эвристические и интеллектуальные методы системного анализа и синтеза // М.: Ленанд. 2015. 306 с.
23. Белов П.Г. Управление рисками, системный анализ и моделирование // Люберцы: Юрайт. 2016. 272 с.
24. Дрогобыцкий И.Н. Системный анализ в экономике: Учебник // М.: ЮНИТИ. 2016. 423 с.
25. Кириллов В.И. Квалиметрия и системный анализ // М.: ИНФРА-М. 2012. 440 с.
26. Козлов В.Н. Системный анализ, оптимизация и принятие решений // М.: Проспект. 2016. 176 с.

R.M.Yusupov, A.A.Musaev (SPIIRAS, Saint-Petersburg)

The problem of evaluating the effectiveness of information technology

The problem of evaluating the effectiveness of information systems and technologies, allowing to estimate the effectiveness of their application is considered. Quantitative assessments of the effectiveness of information technology are proposed. The variants of necessary and sufficient criteria of the effectiveness of information technologies, expressed through the effectiveness of the metasystems in whose interests they function, are presented.

ПРОЕКЦИОННЫЕ ОПЕРАТОРЫ МАТЕМАТИЧЕСКОГО ПРОГРАММИРОВАНИЯ ДЛЯ ИНТЕЛЛЕКТУАЛЬНОЙ ОПТИМИЗАЦИИ УПРАВЛЕНИЙ ДИНАМИЧЕСКИМИ ОБЪЕКТАМИ

В. Н. Козлов, А. А. Ефремов
(*Санкт-Петербургский политехнический университет Петра Великого,*
195251, С.-Петербург, ул. Политехническая, д. 29,
Высшая школа киберфизических систем и управления, saiu@ftk.spbstu.ru)

Введение

Классические методы оптимального управления для синтеза управлений динамическими объектами используют аналитические подходы принципа максимума Л.С. Понтрягина, метода динамического программирования Р. Беллмана и вычислительные методы математического программирования [1]. В докладе развиваются модели интеллектуальных систем и квазианалитические проекционно-операторные методы *математически программируемого интеллектуального управления* для минимизации нормы или линейных функционалов[2 – 5].

Проекционные операторы адекватно описывают обратные связи как решения задач математического программирования. *Класс корректности и универсальности* интеллектуальных систем определяется совместностью обобщенных линейных *ограничений-равенств* задач оптимизации и *ограничения-неравенства* типа нормы, заданного шаром евклидова пространства, а также *совместностью ограничений в целом*.

Операторная форма управлений определяет адекватные разностные операторы систем интеллектуального управления, что создает основу для корректного анализа систем с обратной связью [2].Исследованы достаточные асимптотические свойства разностных операторов интеллектуальных систем методами функционального анализа с учетом реализации заданных требований на примере«программирования многошаговых динамических свойств».

1. Основная часть

Постановка и математическая формулировка задачисинтеза квазиоптимальныхинтервально допустимых систем интеллектуального типа

Требуется вычислить устойчивые интервально допустимые и квазиоптимальные прогнозируемые управления для приближенной минимизации функционала в виде суммы нормы отклонений выходных координат от заданных программных значений и нормы управлений для линейного управляемого по Р. Калману разностного оператора объекта

$$x_{k+1} = Hx_k + F_u u_k \in R^{n_x},\ y_k = cx_k \in R^{n_y},\ u_k \in R^{n_y},\ x_{k_0} = x_0 \in D. \qquad (1)$$

При этом управления формируются на основе решения счетного семейства задач математического программирования вида: вычислить прогнозируемые векторы управлений[2]

$$u_{k*}\left(x_k\right) = \gamma T_U U_{k*}\left(x_k\right),\ \gamma \in R^1,$$

$$U_{k*}\left(x_k\right) = \arg\min\left\{\varphi = \left\|Z_{kp} - C_k\right\|_2^2,\ Z_k = \left(Y_k, U_k\right)^T \,\middle|\, AZ_{kp} = b_k,\ \left\|Z_{kp}\right\|_2^2 \le r^2\right\} \in R^{n_x p}, \qquad (2)$$

на интервале$\left[k, k+p\right] \in N$ как «горизонте прогнозирования» для векторов выходов $Y_k \in R^{n_y p}$ и управлений с учетом ограничений на нормы обобщенного вектора прогнозовZ_k. При этом$D \in R^n -$ область притяжения системы, а векторы и матрицы заданы включениями

$$x_k \in R^{n_x},\ y_k \in R^{n_y},\ H \in R^{n_x \times n_x},\ F \in R^{n_x \times n_u},\ c \in R^{n_y \times n_x}.$$

МатрицаT_U «фильтрует» вектор управления $u_{k*}\left(x_k\right)$ из решения $U_{k*}\left(x_k\right)$задачи (2).

Интеллектуальные свойства системы *программируются в математической форме* путем расширения линейного многообразия ограничений в (2) вида $AZ_{kp} = b_k$. Квазианалитическая проекционно-операторная форма для векторов управлений $U_{k*}\left(x_k\right)$обеспечивает качественный анализ *реализуемых интеллектуальных требований* с помощью обобщенных линейных ограничений, включающих прогнозирующую модель на основе операторов (1) и (2).

В процессе синтеза необходимо исследовать ограничения по устойчивости на параметр обратной связи $\gamma \in R^1$, и качество управления замкнутой нелинейной системы с математическим программированием *интеллектуальных свойств* в ограничениях на координаты состояния и управления. При этом модель объекта управления (1) используется для вычисления прогнозов векторов Y_k в моменты времени $k \in \mathbb{N}$ с учетом целевого вектора C_k, что позволяет вычислить квазиоптимальные допустимые значения векторов выходных координат $y_k \in R^{n_y}$ на интервале прогноза $[k, k+p]$. При этом векторы управлений U_{k*} минимизируют отклонения выходов объекта y_k относительно заданий на текущем интервале.

Синтез интеллектуальных управлений на основе проекционных операторов

Для синтеза определяется структур линейного многообразия $AZ_{kp} = b_k$ с учетом необходимого «интеллектуального обобщения» для математического программирования многошаговой динамики и суммарного ограничения-неравенства на прогнозы векторов выхода и управлений в виде требования $\left\| Z_{kp} \right\|_2^2 \le r^2$. Для совместности ограничений задачи (2)прогнозируемые управления и выходы должны принадлежать *непустому пересечению линейного многообразия и шара*. Пусть система (3) управляема по выходам, т. е. систему можно перевести из одного состояния в другое за конечное время ограниченным управлением. При этом прогнозы выходов и управлений на интервал $[k, k+p] \in N$ задаются обобщенным линейным многообразием с помощью модели типа(1) и«*математически программированием*» прогнозируемой динамики на основе следующего «множества моделей» [2]

$$D_Z^0 = \left\{ Z_{kp} = \left(Y_k, U_k\right)^T \,\middle|\, AZ_{kp} = b_k, \;\; A = \left[E_{pn_y} \,\middle|\, -G \right] \in R^{pn_y \times \left(pn_y + pn_u\right)} \right\}, \qquad (3)$$

Где матрицы $E_{pn_y} \in R^{pn_y \times pn_y}$, $G \in R^{pn_y \times pn_u}$, и векторы Z_{kp} определяют многообразие в виде

$$AZ_{kp} = \begin{bmatrix} E_{n_y} & \dots & 0_{n_y} & \dots & 0_{n_y} & -cF & \dots & 0_{n_y \times n_u} & \dots & 0_{n_y \times n_u} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0_{n_y} & \dots & E_{n_y} & \dots & 0_{n_y} & -cH^{k+j-1}F & \dots & -cF & \dots & 0_{n_y \times n_u} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0_{n_y} & \dots & 0_{n_y} & \dots & E_{n_y} & -cH^{k+p-1}F & \dots & -cH^{k+j-1}F & \dots & -cF \end{bmatrix} \times \begin{bmatrix} y_{k+1} \\ \dots \\ y_{k+p} \\ u_k \\ \dots \\ u_{k+p-1} \end{bmatrix} = c \begin{bmatrix} H \\ \dots \\ H^j \\ \dots \\ H^p \end{bmatrix} x_k = b_k.$$

Тогда обобщенный вектор Z_{kp}, векторы выхода и управлений имеют вид

$$Z_{kp} = \left[Y_k, U_k\right]^T, \;\; Y_k = \left[y_{k+1}, \dots, y_{k+j}, \dots, y_{k+p} \right]^T, \;\; U_k = \left[u_k, \dots, u_{k+j-1}, \dots, u_{k+p-1} \right]^T.$$

Согласованные матричные параметры определены равенствами

$$T_U = \left[0_{pn_y \times pn_u} \,\middle|\, E_{pn_u \times pn_u} \right], \; A = \left[E_{pn_y} \,\middle|\, -G \right], \; rang\, A = pn_y.$$

Блочная матрица и вектор параметров в (3) имеют вид

$$G = \begin{bmatrix} cF & 0_{n_y \times n_u} & \dots & 0_{n_y \times n_u} \\ cHF & cF & \dots & 0_{n_y \times n_u} \\ \dots & \dots & \dots & \dots \\ cH^{p-1}F & cH^{p-2}F & \dots & cF \end{bmatrix} \in R^{pn_y \times pn_u}, \; c\bar{H}x_k = c \begin{bmatrix} H \\ H^2 \\ \dots \\ H^p \end{bmatrix} x_k = b_k.$$

Таким образом, если (H, F, c) управляемы по Р. Калману, то на основе (1) *продолженные решения*(1)из состояния x_k на интервал $[k+1, k+p]$задают прогнозы выходов $y_{k+j}\left(x_k, u_{k+j-1}\right)$ как функции состояния x_k и управлений u_{k+j-1}, $j \in [0, p]$.

Множество (3) как «множеством моделей» и «интеллектуальных целей» задает линейное многообразие для прогнозирования координат и управлений на заданный интервал с помощью продолженных решений системы для выходов. Смешанные ограничения-неравенства на выходы и управления заданым ножеством–параллелепипедом, аппроксимированным шаром

$$D_Z^1 = \left\{ Z_{kp} = \left(Y_k, \mathrm{U}_k\right)^T \;\middle|\; \left\| Z_{kp} \right\|_2^2 \le r^2 \right\}. \qquad (4)$$

Определение решения (3) обобщается для допустимых ограниченных продолженных решений.

Интервально допустимые управления $\bar{U}_k\left(\bar{\theta}_p, x_k\right)$ формируют ограниченные прогнозы выходов динамического объекта на заданном интервале, которые далее вычисляются на основе двух обобщенных «граничных экстремальных решений» задач. *Регуляризованный оператор ИДУ* для целевого функционала по выходу и управлению в функции состояний x_k равен[2]

$$\bar{U}_k\left(\bar{\theta}_p, x_k\right) = T_U\, P_A b_k\left(x_k\right) + \bar{\theta}_p T_U\, P^0 C \rho^{-1/2} \alpha^{1/2}\left(x_k\right),\ \bar{\theta}_p = 1 - 2\, p_\theta(\theta) \in [-1,\ 1],\ \theta \in R, \qquad (4)$$

где два оператора имеют вид $P_A = A^T\left(AA^T\right)^{-1}$, $P^0 = E_{n_y} - P_A A$; кусочно-линейный оператор для $\forall \theta \in R$ равен $p_\theta(\theta) = 0{,}5\left(|\theta| - |\theta - 1| + 1\right) \in [0,\ 1]$; $\rho^{-1/2} = \left\| C^T P^0 C \right\|_2^{-1/2} = \left\| P^0 C \right\|_2^{-1/2} \in R$;

«функциональный» параметр $\alpha^{1/2}\left(x_k\right) = \left[r^2 - p_\alpha\left(\left\| Q_k \right\|_2^2 \right) \right]^{1/2}$, $Q_k = P_A c \bar{H} x_k$, а проектор

$$p_\alpha\left(\left\| Q_k \right\|_2^2 \right) = 0{,}5\left(\left| \left\| Q_k \right\|_2^2 - \varepsilon^2 \right| - \left| \left\| Q_k \right\|_2^2 - (r-\varepsilon)^2 \right| + \varepsilon^2 + (r-\varepsilon)^2 \right) \in \left[\varepsilon^2, (r-\varepsilon)^2 \right] \subset R,$$

является липшицевым: $\left| \alpha^{1/2}(c) - \alpha^{1/2}(d) \right| \le L_\alpha \left| \alpha(c) - \alpha(d) \right|$, где $L_\alpha \in \left[0{,}1r^2;\ 0{,}9r^2 \right]$.

Таким образом, разностный регуляризованный оператор квазиоптимальной системы с учетом (1) и оператора управления с обратной связью (2) определяется уравнениями [2]

$$x_{k+1} = Hx_k + \gamma F_u U_{k*}\left(x_k\right),\ y_k = cx_k,\ x_{k_0} = x_0 \in D,$$

$$U_{k*}\left(x_k\right) = T_U \arg\min\left\{ \varphi = \left\| Z_{kp} - C \right\|_2^2,\ Z = \left(Y_k, U_k\right)^T \;\middle|\; AZ_{kp} = b_k,\ \left\| Z_{kp} \right\|_2^2 \le r^2 \right\} =$$

$$= T_U\, P_A b_k\left(x_k\right) + \bar{\theta}_p T_U\, P^0 C \rho^{-1/2} \alpha_k^{1/2}\left(x_k\right). \qquad (5)$$

позволяет исследовать аналитическими и численными методами устойчивость систем квазиоптимального интервально допустимого управления типа (5). Далее исследованы свойства замкнутых систем, включая их устойчивость и оценки качества процессов.

Аналитические оценки параметров и качества системы по условию устойчивости субоптимальной системы интервального допустимого управления

Исследование устойчивости систем (1), (2) прямым методом Ляпунова связано с построением функции А.М. Ляпунова, что требует дополнительного анализа. Далее приведены достаточные условия устойчивости типа первого приближения на основе принципа сжимающих отображений [1]. Для этого приведены уравнения переходных и стационарных состояний системы, из которых следует уравнение в отклонениях от стационарного состояния как аналог уравнений возмущенного движения по Ляпунову. После этого выполнен переход в левой и правой частях к нормам евклидова пространства. Далее последовательно сформированы уточненные оценки норм правой части, определившие семейство преобразованных неравенств, из которых в силу определения следует условия сжатия оператора (1), (2).Тогда полученные на первом этапе уравнения системы (1), (2) для переходных и стационарных состояний примут вид

$$x_{k+1} = Hx_k + \gamma FT_U\left[P_A c\bar{H}x_k + \bar{\theta}_p P^0 C\rho^{-1/2}\alpha^{1/2}(x_k)\right],\ x_* = Hx_* + \gamma FT_U\left[P_A c\bar{H}x_* + \bar{\theta}_p P^0 C\rho^{-1/2}\alpha^{1/2}(x_*)\right].$$

Для оценки отклонений от стационарного состояния сформировано уравнение

$$x_{k+1} - x_* = H(x_k - x_*) + \gamma FT_U\left[P_A c\bar{H}(x_k - x_*) + \bar{\theta}_p p_C^0\left(\alpha^{1/2}(x_k) - \alpha^{1/2}(x_*)\right)\right],$$

где $p_C^0 = P^0 C\rho^{-1/2}$. Вычисление норм в обеих частях уравнений приводит к равенству

$$\|x_{k+1} - x_*\|_2 = \left\|Hx_k - Hx_* + F\gamma T_U\left[P_A c\bar{H}(x_k - x_*) + \bar{\theta}_p p_C^0\left(\alpha^{1/2}(x_k) - \alpha^{1/2}(x_*)\right)\right]\right\|_2.$$

В силу неравенства треугольника *евклидова норма образанелинейного оператора* правой части последнего равенства преобразуются к форме

$$\|x_{k+1} - x_*\|_2 \le \|H\|_2 \cdot \|x_k - x_*\|_2 + \|F\|_2 \gamma \|T_U\|_2 \times$$
$$\times\left(\|P_A\|_2 \cdot \|c\|_2 \cdot \|\bar{H}\|_2 \cdot \|x_k - x_*\|_2 + \bar{\theta}_p \|p_C^0\|_2 \times \left|\alpha^{1/2}(x_k) - \alpha^{1/2}(x_*)\right|\right). \quad (6)$$

Оценки модуля разности в правой части (6) с учетом матрицы $\bar{D} = \bar{H}^T c^T \bar{A} c\bar{H}$ имеют вид

$$\left|\alpha^{1/2}(x_k) - \alpha^{1/2}(x_*)\right| \le L_\alpha \left|x_k^T \bar{D} x_k - x_*^T \bar{D} x_*\right| \le L_\alpha L_{p\alpha} \|x_k^T + x_*\|_2 \times \|\bar{D}\|_2 \times \|x_k - x_*\|_2, \quad (7)$$

где «ограничивающая функция» с регуляризацией задана равенством

$$\alpha^{1/2}(x_k) = r^2 - p_\alpha\left(\|Q_k\|_2^2\right) = 0{,}5\left(\left|\|Q_k\|_2^2 - \varepsilon^2\right| - \left|\|Q_k\|_2^2 - r^2 - \varepsilon^2\right| + r^2\right),\ \|Q_k\|_2^2 = x_k^T \bar{D} x_k.$$

При этом имеется оценка $\|\bar{D}\|_2 = \|\bar{H}^T c^T \bar{A} c\bar{H}\|_2 \le \|c\|_2^2 \|\bar{H}\|_2^2 \|\bar{A}\|_2$. В результате

$$\left|\alpha^{1/2}(x_k) - \alpha^{1/2}(x_*)\right| \le \|x_k + x_*\|_2 \cdot \|\bar{D}\|_2 \cdot \|x_k - x_*\|_2,\ \|D\|_2 \le \|c\|_2^2 \|\bar{H}\|_2^2 \|\bar{A}\|_2,\ \bar{A} = P_A^T P_A. \quad (8)$$

Оценки в правой части (8) для нормы $\|x_k + x_*\|_2$ вытекают из неравенства треугольника $\|x_k + x_*\|_2 \le \|x_k\|_2 + \|x_*\|_2$ при ограничениях на нормы, имеющих вид

$$\|x_k\|_2^2 \le \bar{\rho} r^2 \beta^{-1},\ \|x_*\|_2^2 \le \bar{\rho} r^2 \beta^{-1},\ \beta = \|c\|_2^2 \|\bar{H}\|_2^2 \|\bar{A}\|_2. \quad (9)$$

Соотношения (9) определены условием корректности вида

$$\alpha^{1/2}(x_k) = r^2 - p_\alpha\left(x_k^T \bar{D} x_k\right) \ge 0,\ \bar{D} = \bar{H}^T c^T P_A^T P_A c\bar{H} = \bar{H}^T c^T \bar{A} c\bar{H}. \quad (10)$$

Условие (10) определяет принадлежность оценок норм выходов допустимой области для прогнозируемых выходов и управлений. Тогда вектор прогнозов выходных координат

$$y_{k+j} = cH^{j-1}Fu_k + ... + cH^{j-2}Fu_{k+j-1} + ... + cFu_{k+j-1} + cH^j x_k,\ j = 1,...,p.$$

Тогда ограничения типа неравенства принимают следующий вид

$$\|Z_{k,p}\|_2^2 \le r^2,\ Z_{k,p} = \left[y_{k+1},...,y_{k+j},...,y_{k+p},u_k,...,u_{k+j-1},...,u_{k+p-1}\right]^T.$$

Квадратичные формы от x_k и x_* в (10) оцениваются по неравенству Коши-Буняковского

$$\left|x_k^T \bar{D} x_k\right| = \left|x_k^T\ \bar{H}^T c^T \bar{A} c\bar{H}\ x_k\right| \rho^{-1} \le \left\|\bar{A}^{1/2} c\bar{H}\right\|_2^2 \|x_k\|_2^2 \le \|c\|_2^2 \|\bar{H}\|_2^2 \|\bar{A}\|_2 \cdot \|x_k\|_2^2 \le r^2. \quad (11)$$

Тогда с учетом (7) – (11) оценку (6)можно преобразовать к окончательному виду

$$\|x_{k+1} - x_*\|_2 \le \|H\|_2 \cdot \|x_k - x_*\|_2 +$$
$$+\|F\|_2 \cdot |\gamma| \cdot \|T_U\|_2 \times \left(\|P_A\|_2 \cdot \|\bar{H}\|_2 \cdot \|c\|_2 + |\bar{\theta}_p| \cdot \|p_C^0\|_2 L_\alpha L_{p\alpha} \|x_k + x_*\|_2 \cdot \|\bar{D}\|_2\right) \|x_k - x_*\|_2,$$

где оценки норм координат и параметров определены неравенствами

$$\| x_k \|_2 \le \rho^{1/2} r \| c \|_2^{-1/2} \| \bar{H} \|_2^{-1/2} \| \bar{A} \|_2^{-1/2}, \quad \| \bar{D} \|_2 \le \rho^{-1} \| c \|_2^2 \| \bar{H} \|_2^2 \| \bar{A} \|_2.$$

Тогдадля нормы в левой части имеет место оценка

$$\| x_{k+1} - x_* \|_2 \le \| H \|_2 \cdot \| x_k - x_* \|_2 + \| F \|_2 \cdot | \gamma | \cdot \| T_U \|_2 \times$$

$$\times \Big(\| P_A \|_2 \cdot \| \bar{H} \|_2 \cdot \| c \|_2 + 2 | \bar{\theta}_p | r \| p_C^0 \|_2 L_\alpha L_{p\alpha} \| \bar{D} \|_2 \Big) \times \| x_k - x_* \|_2, \qquad (12)$$

где использованы оценки $\| x_k \|_2 + \| x_* \|_2 \le 2r \| \bar{D} \|_2$. Тогда в силу оценки

$$r \| \bar{D} \|_2 = r \| c \|_2^{-1/2} \| \bar{H} \|_2^{-1/2} \| \bar{A} \|_2^{-1/2} \le r \left(\| c \|_2 \cdot \| \bar{H} \|_2 \right)^{-1/2} \| \bar{A} \|_2^{-1/2},$$

из неравенства (12) следует *условие сжатия* для оператора интеллектуальной системы [2]

$$\| x_{k+1} - x_* \|_2 \le s \| x_k - x_* \|_2,$$

$$s = \| H \|_2 + \| F \|_2 \cdot |\gamma| \cdot \| T_U \|_2 \times \Big(\| P_A \|_2 \cdot \| \bar{H} \|_2 \cdot \| c \|_2 + 2r |\bar{\theta}_p| \cdot \| p_C^0 \|_2 L_\alpha L_{p\alpha} \| \bar{D} \|_2 \Big) < 1. \quad (13)$$

Из этого неравенства следует ограничение на параметр обратной связи на основе достаточного условие устойчивости в форме условия сжатия разностного оператора системы п-оптимального интервально допустимого управления, которое имеет вид

$$| \gamma | < \left(1 - \| H \|_2 \right) \times \Big(\| P_A \|_2 \cdot \| \bar{H} \|_2 \cdot \| c \|_2 + 2 | \bar{\theta}_p | r \| p_C^0 \|_2 L_\alpha L_{p\alpha} \| \bar{D} \|_2 \Big)^{-1} \Big(\| F \|_2 \cdot \| T_U \|_2 \Big)^{-1}. \quad (14)$$ Нер

авенство (13) с учетом параметра $s \in R$ определяет условие сжатия для функций Ляпунова

$$V_{k+1} \le s V_k, \quad V_k \triangleq \| x_k - x_* \|_2,$$

$$s = \| H \|_2 + \| F \|_2 \cdot |\gamma| \cdot \| T_U \|_2 \times \Big(\| P_A \|_2 \cdot \| \bar{H} \|_2 \cdot \| c \|_2 + 2r |\bar{\theta}_p| \cdot \| p_C^0 \|_2 L_\alpha L_{p\alpha} \| \bar{D} \|_2 \Big) < 1. \quad (15)$$

Оценки параметров (13) – (15) по условию устойчивости являются достаточными, однако определяют непустое множество стабилизирующих управлений типа (2), что далее подтверждено вычислительными экспериментами для анализа устойчивости и качества интеллектуальных систем.

2. Оценки устойчивости и качества интеллектуальных систем

В качестве объектов анализа рассмотрены объекты и системы первого и второго порядков, для которых вычислены интервально допустимые прогнозируемые управления, сформированы операторы замкнутых систем, исследованы устойчивость и качество в переходных и асимптотических режимах, а также. расширенные требования к устойчивости по параметру γ.

Эксперимент 1 посвящен исследованию субоптимальной системы первого порядка с прогнозом на $p-$интервал. Дан анализ устойчивости и качества для одномерного объекта

$$x_{k+1} = H x_k + F_u u_k,\ y_k = c x_k,\ x_{k_0} = x_0,\ H = 0.5,\ F = 2,\ c = 1.5,\ x_0 = 2. \quad (16.\text{а})$$

Требовалось исследовать устойчивость системы на основе вычисления прогнозов управлений на дискретном временном интервале для $k \in [0,4]$, заданных равенством

$$U_{k*} = T_U \arg\min \left\{ \varphi = \| Z_{kp} - C \|_2^2,\ Z = \left(Y_k, U_k \right)^T \,\middle|\, A Z_{kp} = b_k,\ \| Z_{kp} \|_2^2 \le r^2 \right\}, \quad (16.\text{б})$$

где прогнозируемый вектор «выходов-управлений» и параметры имеют вид

$$Z_{kp} = \left(y_1 \quad y_2 \quad y_3 \quad y_4 \quad y_5 \quad u_0 \quad u_1 \quad u_2 \quad u_3 \quad u_4 \right)^T,$$

$$r = 2,\ C = \left(1 \quad 0.5 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \right)^T.$$

Синтез системы п-оптимального скалярного интервально допустимого управления выполнен на основе«обобщенной модели», заданной линейным многообразием типа (3), для которого функциональные и числовые параметры определены соотношениями

$$z_k \triangleq \begin{bmatrix} y_1 & y_2 & y_3 & y_4 & y_5 & u_0 & u_1 & u_2 & u_3 & u_4 \end{bmatrix}^T.$$

В результате числовые матрица A, вектор b_k и«фильтрующая» матрица T_u равны

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -3 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1.5 & -3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -0.75 & -1.5 & -3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -0.38 & -0.75 & -1.5 & -3 & 0 \\ 0 & 0 & 0 & 0 & 1 & -0.19 & -0.38 & -0.75 & -1.5 & -3 \end{bmatrix}, \; b_k = \begin{bmatrix} 1.5 \\ 0.75 \\ 0.38 \\ 0.19 \\ 0.09 \end{bmatrix}, \; T_U = \begin{bmatrix} 0_5 & E_5 \end{bmatrix}.$$

Интервально допустимое проекционное управление типа (5), имеющее вид

$$U_{k*} == T_U P_A b_k(x_0) + \overline{\theta}_p T_U P^0 C \rho^{-1/2} \alpha_k^{1/2}(x_0), \qquad (17)$$

определено числовыми матрицами, заданными включениями и равенствами

$$P_A = A^T\left(AA^T\right)^{-1} \in R^{10\times 5}, \; P^0 = E_{10} - P_A A \in R^{10\times 10}, \; \overline{\theta}_p = -0.1, \;\; \rho^{-1/2} = 0.94, \;\; \alpha^{1/2}(x_0) = 1.97.$$

Таким образом, интервально допустимые управления равны

$$U_{k*} = \begin{bmatrix} u_0 & u_1 & u_2 & u_3 & u_4 \end{bmatrix}^T = \begin{bmatrix} -0.51 & -0.02 & -0.01 & -0.01 & 0 \end{bmatrix}^T.$$

На рис. 1 иллюстрируется пример интервально допустимого управления $U_{k*} \in R^1, \; r = 2.$

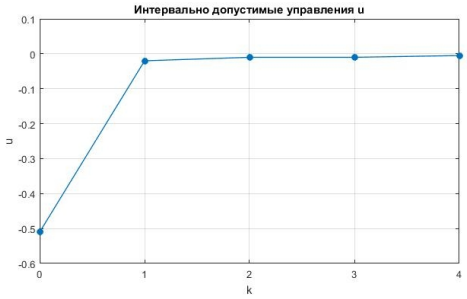


Рис. 1. Пример субоптимального интервального допустимого управления

Устойчивость системы исследована для модели (1), (2), обобщенной в (5) к виду

$$x_{k+1} = Hx_k + \gamma F T_U\left(P_A b_k(x_0) + \overline{\theta}_p P^0 C \rho^{-1/2} \alpha^{1/2}(x_0)\right), \; y_k = cx_k, \; x_{k_0} = x_0. \quad (18)$$

Устойчивость системы (5) гарантирована ограничением на параметр обратной связи γ вида

$$|\gamma| < \left(1 - \|H\|_2\right) \times \left(\|P_A\|_2 \cdot \|\overline{H}\|_2 \cdot \|c\|_2 + 2|\overline{\theta}_p| r \left\|P^0 C\right\|_2 L_\alpha L_{p\alpha} \|\overline{D}\|_2\right)^{-1} \left(\|F\|_2 \cdot \|T_U\|_2\right)^{-1},$$

где $L_\alpha, L_{p\alpha} \in [0.4;\; 3.6]$ - постоянные Липшица, матрица $\overline{D} = \overline{H}^T c^T P_A^T P_A c \overline{H}$.Вычисление оценок для системы (16) определяет ограничение на параметр обратной связи: $|\gamma| < 0.62.$

Таким образом, ограничение параметра γ определяет такой сжимающий разностный оператор, для которого $V_{k+1} \le s(\gamma) V_k, \; V_k \triangleq \|x_k - x_*\|_2, \; s(\gamma) < 1,$ для системы (16), (17).Далее вычислены прогнозы выходов и состояний системы на интервал $k \in [0,4]$ при различных значениях γ для вектора *целевых значений выходов* системы

$$C_k = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 & y_5 \end{bmatrix}^T = \begin{bmatrix} 1 & 0.5 & 0 & 0 & 0 \end{bmatrix}^T.$$

Анализ динамики асимптотически устойчивого разностного оператора (16),(17) выполнен при $\gamma = 0.5.$ Вычислены прогнозы выходов на начальной стадии процесса $k \in [0,\; 4]$:

$$x_{k+p} = [x_{k+0} \quad x_{k+1} \quad x_{k+2} \quad x_{k+3} \quad x_{k+4}]^T = [0.49 \quad 0.22 \quad 0.01 \quad 0.01 \quad 0]^T,$$
$$y_{k+p} = [y_{k+0} \quad y_{k+1} \quad y_{k+2} \quad y_{k+3} \quad y_{k+4}]^T = [0.74 \quad 0.33 \quad 0.02 \quad 0.01 \quad 0].$$

На рис. 2 приведены прогнозы выхода y_{k+1} на заданный интервал (сплошная линия) и целевые значения $C_k \in R^1$ выходных координат на интервале управления (пунктирная линия).

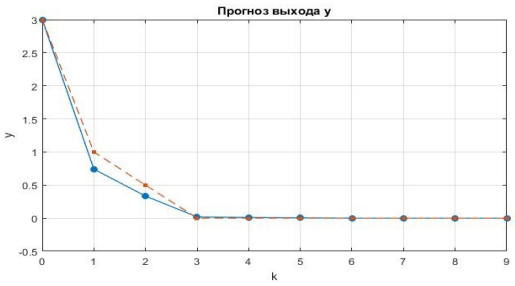


Рис. 2. Процесс изменения выходной координаты y_k в системе ($\gamma = 0.5$)

Эксперимент подтверждает существование непустого множества устойчивых интеллектуальных систем, в частности, данные рис. 2 следует, поскольку система для $\gamma = 0.5$ устойчива. Проекционно-операторное допустимое управление дает корректные прогнозы выходов системы (1), (2) на заданный интервал. Устойчивость процессов в системе, при параметре $\gamma = 3$, *не удовлетворяющем аналитической оценке* (14), отражает *достаточность* оценок (13) – (15), следующих из условий сжатия оператора системы (1), (2) т. к. это условие является «сильно» достаточным. Состояния и прогнозы выходов для $k \in [0, \ 5]$ имеют вид

$$x = [x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5]^T = [-2.04 \quad -1.16 \quad -0.66 \quad -0.32 \quad -0.28]^T,$$
$$y = [y_1 \quad y_2 \quad y_3 \quad y_4 \quad y_5]^T = [-3.06 \quad -1.75 \quad -0.98 \quad -0.49 \quad -0.42]^T.$$

На рис. 3 дана *начальная часть* асимптотического процесса по выходу y_{k+1} (сплошная линия),

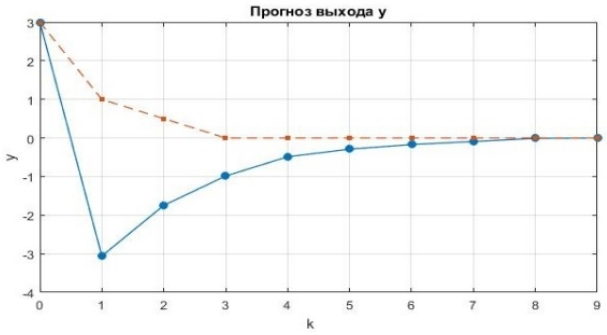


Рис. 3. Переходная часть асимптотического процесса по выходу y_k системы $(\gamma = 3)$

а его целевые значения C_k (пунктир) иллюстрируют устойчивость и снижение качества.

Эксперимент 2. Исследованы асимптотическая устойчивость и качество системы с прогнозом трех состояний $(p = 3)$ для модели (16.а), (16.б)с объектом 2 порядка с параметрами

$$H = \begin{bmatrix} 0.5 & 0.1 \\ 0 & 0.5 \end{bmatrix}, \quad F = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}, \quad c = \begin{bmatrix} 0.8 & 0 \\ 0 & 1 \end{bmatrix}, \quad x_0 = \begin{bmatrix} 6 \\ 4 \end{bmatrix},$$

при начальных условиях двух типов: 1) $x_0 = \pm 1$; 2) $x_0 = \pm 10$; для $\gamma = 0.5$ в (13). Исследование подтвердило устойчивость и качество интеллектуальной системы при прогнозировании асимптотической динамики системы второго порядка, а также выводы о дополнительной оптимизации системы по параметру $\gamma = 0.5$. Прогноз выходов и управлений

$$U_{k*} = T_U \arg\min \left\{ \varphi = \left\| Z_{kp} - C \right\|_2^2, \ Z = (Y_k, U_k)^T \ \middle| \ AZ_{kp} = b_k, \ \left\| Z_{kp} \right\|_2^2 \le r^2 \right\},$$

где расширенный вектор состояния имеет структуру

$$Z_{kp} = \left(y_1^1 \quad y_1^2 \quad y_2^1 \quad y_2^2 \quad y_3^1 \quad y_3^2 \quad y_4^1 \quad y_4^2 \quad u_0^1 \quad u_0^2 \quad u_1^1 \quad u_1^2 \quad u_2^1 \quad u_2^2 \quad u_3^1 \quad u_3^2 \right)^T,$$

а радиус шара и целевой вектор имеют вид: $r=3,\ C_k = (1.5 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)^T$.

В многообразии (3) вектор $z_k = \left[y_1 \quad y_2 \quad y_3 \quad y_4 \quad u_0 \quad u_1 \quad u_2 \quad u_3 \right]^T$, а параметры равны:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1.2 & -0.16 & -2.4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -0.6 & -0.16 & -1.2 & -0.16 & -2.4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -0.5 & 0 & -1 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -0.3 & -0.12 & -0.6 & -0.16 & -1.2 & -0.16 & -2.4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -0.25 & 0 & -0.5 & 0 & -1 & 0 & -2 \end{bmatrix}, \quad b_k = \begin{bmatrix} 2.72 \\ 2 \\ 1.52 \\ 1 \\ 0.84 \\ 0.84 \\ 0.46 \\ 0.25 \end{bmatrix}.$$

Как и в задаче для объекта первого порядка, интервально допустимое управление, вычисленное с помощью проекционно-операторного метода, имеет вид

$$U_{k*} = T_U P_A b_k \left(x_0 \right) + \bar{\theta}_p T_U P^0 C \rho^{-1/2} \alpha^{1/2} \left(x_0 \right),$$

где $P_A = A^T \left(AA^T \right)^{-1} \in R^{16\times 8}$, $P^0 = E_{16} - P_A A \in R^{16\times 16}$, $\bar{\theta}_p = -0.1,\ \rho^{-1/2} = 0.62,\ \alpha^{1/2}\left(x_0 \right) = 2.84.$

Интервально допустимые управления на начальной стадии процесса определены равенствами

$$U_{k*} = \left[u_0^1 \quad u_0^2 \quad u_1^1 \quad u_1^2 \quad u_2^1 \quad u_2^2 \quad u_3^1 \quad u_3^2 \right]^T = \left[-1.06 \quad -0.88 \quad -0.04 \quad -0.05 \quad 0.01 \quad -0.01 \quad 0 \quad 0 \right]^T.$$

На рис. 4 изображены графики вычисленных интервально допустимых управлений $U_{k*} \in R^2$.

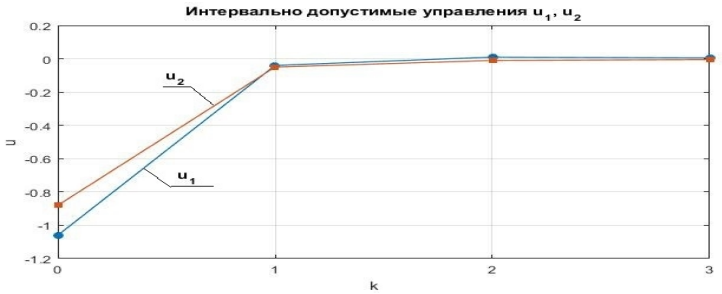


Рис. 4. Интервально допустимые управления на начальной переходной стадии процесса

Можно отметить, что управление на каждом шаге удовлетворяет ограничениям задачи и находится внутри шара радиусом $r = 3$. Исследование устойчивости выполнено для системы (1), (2) в форме (5) с учетом проекционного квази оптимального управления, имеющего вид

$$x_{k+1} = Hx_k + \gamma F T_U \left(P_A b_k \left(x_0 \right) + \bar{\theta}_p P^0 C \rho^{-1/2} \alpha^{1/2} \left(x_0 \right) \right),\ y_k = cx_k,\ x_{k_0} = x_0. \qquad (20)$$

Как и в ранее, устойчивость системы (20) зависит от параметра γ, имеющего смысл коэффициента усиления обратной связи, для которого имеет место достаточная оценка (14). Вычислив правую часть в неравенстве (14) с учетом соотношений где $L_\alpha, L_{p\alpha} \in [0.9;\ 8.1]$ - постоянные Липшица, $\bar{D} = \bar{H}^T c^T P_A^T P_A c \bar{H}$, можно получить ограничение на коэффициент усиления обратной связи в виде неравенства $|\gamma| < 0.37.$

Таким образом, параметр γ со значением, удовлетворяющим условию устойчивости системы (20), определяет отображение $V_{k+1} \le s(\gamma) V_k, \quad V_k \triangleq \| x_k - x_* \|_2,\ s(\gamma) < 1.$

Далее вычислены прогнозы выходных координат и состояний системы (20) на интервале $k \in [0,\ 3] \in \mathbb{Z}$ при различных значениях параметра обратной связи γ и векторах целевых значений выходов C_k, которые имеют вид

$$C_k = [y_1^1 \quad y_1^2 \quad y_2^1 \quad y_2^2 \quad y_3^1 \quad y_3^2 \quad y_4^1 \quad y_4^2]^T = [1.5 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]^T.$$

Вычислительный эксперимент включает исследование динамики процессов системы (20) для параметра $\gamma = 0.3$, удовлетворяющего условию устойчивости замкнутой системы. *Цели эксперимента заключались* **в** анализе корректности вычисления прогнозов выходов и управлений как функций состояний замкнутой системы (20) при $\gamma = 0.3$ и обеспечении асимптотической устойчивости системы. Значения прогнозируемых выходов и состояний для (20) на начальном (переходном) этапе процесса управления определены равенствами

$$y_k = [y_1^1 \quad y_1^2 \quad y_2^1 \quad y_2^2 \quad y_3^1 \quad y_3^2 \quad y_4^1 \quad y_4^2]^T = [1.96 \quad 1.47 \quad 0.01 \quad 0.03 \quad 0 \quad 0.01 \quad 0 \quad 0]^T,$$

$$x_k = [x_1^1 \quad x_1^2 \quad x_2^1 \quad x_2^2 \quad x_3^1 \quad x_3^2 \quad x_4^1 \quad x_4^2]^T = [2.44 \quad 1.47 \quad 0.01 \quad 0.03 \quad 0.01 \quad 0.01 \quad 0 \quad 0]^T.$$

На рис. 5иллюстрируется асимптотическая устойчивость процессов в системе и близость в начальном (переходном) интервале времени прогнозов выходов y_{k+1}^1, y_{k+1}^2(сплошная линия) и целевых значений C_k^1, C_k^2 для соответствующих выходов $y^1,\ y^2$ (пунктирная линия).

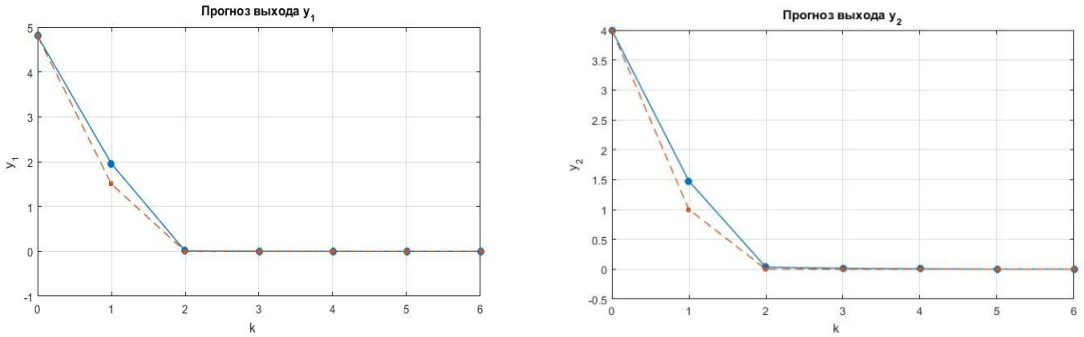


Рис. 5. Процессы изменения выходных координат y_k^1, y_k^2 системы $(\gamma = 0.3)$

Из данных результатов следует, что проекционные операторы управления позволяют синтезировать асимптотически устойчивые системы интеллектуального управления.

3. Заключение

Математическая формализация «интеллектуальных свойств и заданных требований» реализуется явной формой операторно-оптимальных обратных связей с законами управления, обеспечивающими интеллектуализацию систем на основе универсальной корректных математических формулировок задач синтеза. Сформулированы *основы синтеза* интеллектуальных систем с динамическими или статическими объектами управления с математически программируемыми интеллектуальными свойствами:

- системы локально-допустимого и локально-оптимального управления;
- системы интервально-допустимого и интервально-оптимального управления;

Для указанных классов систем на основе принципа сжимающих отображений исследованы условия устойчивости, определяющие требования к параметрам статической обратной связи.

Предлагаемые результаты использованы для решения следующих задач:

- управление частотой и перетоками активной мощности по линиям сверхбольших энергетических объединений;
- управление транспортировкой жидких углеводородов по магистральным трубопроводам систем энергоснабжения.

ЛИТЕРАТУРА

1. Kirk D. Optimal Control Theorie and Introduction. Mineola, New York. 1998. 452 p.
2. Козлов В. Н. Негладкие системы, операторы оптимизации и устойчивость энергообъединений. СПб.: Изд-во Политехн. ун-та. 2012. -183 с.
3. Козлов В. Н. Устойчивость динамических систем с операторами минимизации. Сборник трудов 5 Четаевской конференции. Казань. Изд-во Казанского национ. иссл. ун-та (КАИ). 2012.
4. Козлов В. Н. Синтез управлений крупномасштабными объектами на основе операторов оптимизации // Материалы 6-й Всероссийской мультиконференции «Проблемы управления», т. 3, Изд-во Южного федерального ун-та, Ростов-на-Дону. 2013. С. 195–197.
5. Kozlov V. N. The Metod of Minimization of Linear Functionals Based on Compakt Sets. Proceedings of the 12^{th} international workshop on computer science and information technologies (CSIT 2010), Russia, Moskov-Saint-Petersburg, Russia, 2010.Volume 2. pp. 157–159.

ПОСТРОЕНИЕ МНОГОУРОВНЕВОЙ БАЗЫ ДЛЯ УМЕНЬШЕНИЯ ВЫЧИСЛИТЕЛЬНОЙ СЛОЖНОСТИ РЕШЕНИЯ ЗАДАЧИ КОНЪЮНКТИВНЫЙ БУЛЕВСКИЙ ЗАПРОС

Т. М. Косовская[1]
(*Санкт-Петербургский государственный университет,*
199034, С.-Петербург, Университетская наб., 7/9,
kosovtm@gmail.com)

Аннотация

В докладе сравниваются две NP-полные задачи КОНЪЮНКТИВНЫЙ БУЛЕВСКИЙ ЗАПРОС (КБЗ), касающаяся баз данных, и ВЫПОЛНИМОСТЬ В КОНЕЧНОЙ ИНТЕРПРЕТАЦИИ (ВКИ), касающаяся решения задач искусственного интеллекта, имеющие практически одинаковые формулировки, но имеющие существенно различное применение. Описывается построение многоуровневой базы данных, позволяющей уменьшить время решения задачи КОНЪЮНКТИВНЫЙ БУЛЕВСКИЙ ЗАПРОС. Описывается алгоритм выделения множеств литералов, конъюнкции которых изоморфны друг другу.

Ключевые слова: конъюнктивный булевский запрос, вычислительная сложность алгоритмов.

Введение

В современных научных исследованиях большое внимание уделяется анализу алгоритмов, решающих NP-полные задачи. В некоторых случаях удаётся свести задачу к последовательному решению серии аналогичных задач с исходными данными меньшей длины. В докладе предлагается алгоритм построения многоуровневой базы данных, обеспечивающей возможность уменьшения вычислительной сложности NP-полной задачи КОНЪЮНКТИВНЫЙ БУЛЕВСКИЙ ЗАПРОС (КБЗ) [1] при её многократном решении.

КОНЪЮНКТИВНЫЙ БУЛЕВСКИЙ ЗАПРОС (КБЗ)

Дано. Конечное множество D; совокупность предикатов $R = \{R_1, \ldots, R_n\}$, где R_i задаёт d_i-местное отношение между элементами из D;

множество $S(D)$ всех литералов с предикатами из R, истинных на D;

конъюнктивный булевский запрос Q с литералами вида $R_j(u_1, \ldots, u_{d_j})$ с переменными и константами из D в качестве аргументов.

Вопрос. Существует ли набор значений из D для переменных, при которых Q истинен?

$$S(D) \Rightarrow \exists y_1 y_2 \ldots y_l (A_1 \,\&\, A_2 \,\&\, \ldots \,\&\, A_r).$$

Формулировка этой задачи почти полностью совпадает с NP-полной задачей выделения объекта на сложной сцене при логико-предметном подходе к решению задач искусственного интеллекта.

ВЫПОЛНИМОСТЬ В КОНЕЧНОЙ ИНТЕРПРЕТАЦИИ (ВКИ)

ДАНО. Множество $\omega = \{\omega_1, \ldots \omega_t\}$, набор предикатов $\{p_1, \ldots, p_n\}$, задающих свойства элементов из ω и отношения между ними;

набор $S(\omega)$ истинных постоянных литералов вида $p_i(\boldsymbol{\tau}_{ij})$, где $i = 1, \ldots, n$, $\boldsymbol{\tau}_{ij}$ – упорядоченные списки элементов из ω;

бескванторная формула $A(\boldsymbol{y})$, представленная в виде дизъюнкции элементарных конъюнкций атомарных формул ($\boldsymbol{y} = (y_1, \ldots, y_a)$ – список всех предметных переменных, входящих в формулу $A(\boldsymbol{y})$).

ВОПРОС. Существует ли набор значений для $\boldsymbol{y}$ из ω^a, для которого формула $A(\boldsymbol{y})$ истинна?

$$S(\omega) \Rightarrow \exists \boldsymbol{y}\ \mathrm{A}(\boldsymbol{y}).$$

Существенная разница в применении этих задач заключается в следующем:

[1] Доктор физико-математических наук, профессор.

– база данных может не изменяться вообще или меняться достаточно редко, а запросы каждый раз могут возникать различные;
– при распознавании образов множество целевых формул (описаний классов объектов) может не изменяться вообще или меняться достаточно редко, а распознаваемые объекты каждый раз могут возникать новые.

В докладе сначала приводятся основные оценки сложности решения задач логико предметного распознавания как при их непосредственном решении, так и при построении многоуровневого описания классов, совпадающие с таковыми при решении задачи КБЗ.

Основным результатом доклада являются алгоритмы построения многоуровневой базы данных, существенно отличающиеся от алгоритмов построения многоуровневого описания классов.

Задачи логико-предметного распознавания

Для решения задачи распознавания образов автором ранее было предложено создание многоуровневого описания классов [2, 3]. Такое описание основано на выделении из совокупности элементарных конъюнкций, составляющих описание класса, формул, изоморфных часто встречающимся подформулам этих элементарных конъюнкций, имеющим небольшую сложность.

Определение. *Две элементарные конъюнкции атомарных формул исчисления предикатов P и Q называются изоморфными, если существуют такая элементарная конъюнкция R и подстановки переменных из формулы R вместо всех вхождений переменных в формулы P и Q, что результаты этих подстановок совпадают с формулой R с точностью до порядка литералов.*

При этом полученные подстановки $(x_1 \to a_1, \ldots, x_n \to a_n)$ *и* $(x_1 \to b_1, \ldots, x_n \to b_n)$ *называются унификаторами формул P и Q с формулой R соответственно.*

В [4] доказаны оценки числа шагов алгоритмов, решающих задачу ВКИ. Так, например, для алгоритма полного перебора всех возможных подстановок констант из $\omega = \{\omega_1, \ldots, \omega_t\}$ вместо переменных из $A(\boldsymbol{x})$ число шагов составляет $O(t^m)$, где t – число элементов в ω, m – количество переменных в $A(\boldsymbol{x})$. Для алгоритмов, основанных на построении вывода в исчислении предикатов, оценки имеют вид $O(s^a)$, где s – максимальное количество вхождений одинаковых предикатных символов в $S(\omega)$, a – суммарное количество вхождений предикатных символов в $A(\boldsymbol{x})$.

Отметим, что оценки числа шагов поиска ответа на запрос при решении задачи КБЗ такие же, как при решении задачи ВКИ.

Для уменьшения показателя степени верхней оценки числа шагов алгоритма, решающего эту задачу, в [2] было предложено понятие многоуровневой системы описания целевых условий, учитывающее «часто» встречающиеся изоморфные подформулы «небольшой сложности». Были также доказаны достаточные условия уменьшения числа шагов для двух видов алгоритмов, уточняющие для каждого из них нечёткие понятия «часто» и «небольшой сложности».

Петровым Д.А. разработана программа, осуществляющая выделение максимальной формулы, изоморфной подформулам двух элементарных конъюнкций [5], а также программа построения многоуровневого описания класса.

Построение многоуровневой базы данных для решения задачи КБЗ

При многократном решении задачи КБЗ в распоряжении исследователя нет совокупности элементарных конъюнкций. Аналоги часто встречающихся подформул приходится выделять из совокупности литералов, и алгоритм построения многоуровневого описания классов не может быть применён. При создании базы данных мы не можем сказать с уверенностью, какие запросы могут возникнуть у пользователя. Однако, сама база данных при всех запросах остаётся практически неизменной. Поэтому закономерности следует искать в самой базе (множестве постоянных литералов $S(D)$).

Алгоритм нахождения таких закономерностей (множеств атомарных формул конъюнкции которых изоморфны друг другу) приведён ниже. В этом разделе рассмотрим построение многоуровневой базы данных и изменение числа шагов поиска ответа на запрос при её использовании.

Построение 2-уровневой базы данных

1. Пусть в множестве атомарных формул, входящих в S(D) выделены подмножества, в каждом из которых их конъюнкция изоморфна формуле $P_i^1(\boldsymbol{y}_i^1)$ ($i = 1, \ldots, n_1$) и для каждой группы найден общий унификатор конъюнкции ее элементов с $P_i^1(\boldsymbol{y}_i^1)$.

2. Введём предикат 1-го уровня $p_i^1(y_i^1)$, определяемый равносильностью $p_i^1(y_i^1) \Leftrightarrow P_i^1(\boldsymbol{y}_i^1)$, где y_i^1 – переменные первого уровня для списков исходных переменных $\boldsymbol{y}_i^1$.

3. Добавим в $S(D)$ множество постоянных атомарных формул вида $p_i^1(d_i^{\{1,j\}})$, где $d_i^{\{1,j\}}$ – обозначение для списка констант из D, для которых выполняется $P_i^1(\boldsymbol{y}_i^1)$. В результате получаем 2-уровневую базу $S^1(D)$.

Заметим, что построение 2-уровневой базы данных – NP-трудная задача с огромными исходными данными, так как в п. 1 решается NP-трудная задача выделения из всех формул базы подмножеств небольшой мощности, в каждом из которых их конъюнкция изоморфна одной и той же формуле.

Эта задача решается один раз.

Применение двухуровневой базы данных

Пусть получен конъюнктивный булевский запрос $\exists y_1\, y_2 \ldots y_l\, A(y_1,y_2,\ldots y_l, d_1, \ldots, d_r)$, где $d_1, \ldots, d_r$ – константы из D.

1. Проверка логического следования

$$A(y_1,y_2,\ldots y_l, d_1, \ldots, d_r) \Rightarrow \exists y_1\, y_2 \ldots y_l\, P_i^1(\boldsymbol{y}_i^1)$$

в случае, если оно выполняется, позволяет найти все подформулы запроса, изоморфные $P_i^1(\boldsymbol{y}_i^1)$, и их унификаторы.

Несмотря на то, что задача NP-трудна, ее исходные данные имеют небольшую длину и оценки числа шагов имеют в показателе степени параметры формулы $P_i^1(\boldsymbol{y}_i^1)$ (количество переменных или количество атомарных формул с одним и тем же предикатом). Причем эти параметры меньше, чем соответствующие параметры формулы $A(y_1,y_2,\ldots y_l, d_1, \ldots, d_r)$.

2. Заменим в элементарной конъюнкции $A(y_1,y_2,\ldots y_l, d_1, \ldots, d_r)$ все найденные подформулы, изоморфные $P_i^1(\boldsymbol{y}_i^1)$ на $p_i^1(y_i^1)$, где y_i^1 – переменная для списка исходных переменных и констант этой подформулы. Это возможно в силу наличия унификатора для каждой подформулы элементарной конъюнкции с изоморфной ей. Получим элементарную конъюнкцию $A^1(\boldsymbol{x}^1)$, где $\boldsymbol{x}^1$ – список переменных (как основных, так и 1-го уровня) и констант, оставшихся в явном виде в качестве аргументов.

Этот пункт выполняется за линейное от длины записи $A(y_1,y_2,\ldots y_l, d_1, \ldots, d_r)$ число шагов. Заметим, что количество атомарных формул в $A^1(\boldsymbol{x}^1)$ не превосходит количества атомарных формул в $A(y_1,y_2,\ldots y_l, d_1, \ldots, d_r)$ (и строго меньше в случае, если была заменена хоть одна подформула на атомарную 1-го уровня).

3. При поверке того, что $S^1(D) \Rightarrow \exists\, \boldsymbol{y}_i^1\ A^1(\boldsymbol{y}_i^1)$ в первую очередь проверяем это логическое следование для атомарных формул с предикатными символами 1-го уровня.

Это потребует

$O(t_1^{\{l1\}})$ шагов при использовании переборного алгоритма, где t_1 – количество элементов в базе $S^1(D)$ с предикатами 1-го уровня, где $l1$ – количество переменных 1-го уровня в запросе и

$O(\Sigma\{i:\ p_i^1\ is\ in\ A^1(\boldsymbol{x}^1)\}s\{i,1\}^{a\{i,1\}})$, при использовании логических алгоритмов, где $s\{i,1\}$ и – количество вхождений предиката p_i^1 в описание $S^1(D)$ и формулу $A^1(\boldsymbol{x}^1)$ соответственно.

Очевидно, что количество шагов проверки выполнения булевского запроса уменьшится.

Нахождение множеств атомарных формул, конъюнкции которых изоморфны друг другу

Описанный ниже алгоритм предложен и реализован в дипломной работе Глазырина П.А., выполненной под моим руководством, и имеет дело именно с совокупностью литералов. Его идеи заимствованы из [6], где выделяются наибольшие общие подформулы из пропозициональной формулы в ДНФ (дизъюнктивная нормальная форма). Существенным отличием излагаемого здесь алгоритма от предложенного в [6], является то, что из формулы в

ДНФ выделяются не совпадающие друг с другом пропозициональные формулы, а конъюнкции предикатных формул, изоморфные друг другу.

Для дальнейшего изложения потребуются некоторые определения и теорема.

Определение. *Пара – это конъюнкция литералов, упорядоченная по номеру входящих в них литералов, причём если предикатные символы в паре одинаковы, то сначала находится литерал без отрицания (если он есть), а затем литерал с отрицанием (если он есть).*

***Определение**. Характеристикой пары называется две последовательности, полученных следующим образом:*

первая последовательность получается упорядочиванием всех аргументов пары в порядке количества их вхождений в пару, причём аргументы с одинаковым количеством вхождений упорядочиваются в порядке их появления;

вторая последовательность получается выписыванием номеров позиций, на которых стоят аргументы в первой паре.

Например, характеристикой пары $\neg p_i(a,b.c.a)$ & $p_j(d,a,c,b,u)$ являются две последовательности *(d,u,b,b,c,c,a,a,a)* и *(5,9,2,8,3,7,1,4,6)*.

Теорема. Д*ля того, чтобы две пары литералов были изоморфны необходимо и достаточно, чтобы*

– их предикатные символы совпадали и литералы с одинаковыми предикатными символами в разных парах были одновременно без отрицания или одновременно с отрицаниями;

– вторые последовательности их характеристик совпадали, если предикатные символы в паре различны или предикатные символы в паре одинаковы, но один из литералов без отрицания, а другой с отрицанием;

– вторая последовательность характеристики первой пары совпадает либо со второй последовательностью характеристикой второй пары, либо со второй последовательностью характеристикой второй пары с переставленными литералами.

Д о к а з а т е л ь с т в о. Н е о б х о д и м о с т ь.

Необходимость первого условия очевидна.

Пусть имеются две пары $A = p_i^{\alpha}(a_1, \ldots, a_k)$ & $p_j^{\beta}(a_{k+1}. \ldots, a_m)$ и $B = p_i^{\alpha}(b_1. \ldots, b_k)$ & $p_j^{\beta}(b_{k+1}, \ldots, b_m)$ ($i \neq j$ или $i = j$ при $\alpha \neq \beta$; верхние индексы у предикатных символов указывают на наличие или отсутствие отрицания у литералов) и первые последовательности их характеристик равны соответственно $(a'_1. \ldots, a'_m)$ и $(b'_1. \ldots, b'_m)$.

Покажем, что эти пары изоморфны, то есть существует элементарная конъюнкция $C= p_i(x_1. \ldots, x_k)$ & $p_j(x_{k+1}, \ldots, x_m)$ (x_{i1} может совпадать с x_{i2} при $i1 \neq i2$) и такие подстановки переменных $x_1, \ldots, x_k, x_{k+1}, \ldots, x_m$ вместо констант $a_1, \ldots, a_k, a_{k+1}. \ldots, a_m$ и $b_1. \ldots, b_k, b_{k+1}, \ldots, b_m$ соответственно, что результаты этих подстановок совпадают с C.

Последовательно каждый аргумент первой пары заменим на переменную, при этом разным аргументам пары соответствуют разные переменные, а одна и та же константа заменяется на одну и ту же переменную. Унификатор A и C будет иметь вид $(x_1 \rightarrow a'_1, \ldots, x_n \rightarrow a'_m)$.

Аналогично унификатор B и C будет иметь вид $(x_1 \rightarrow b'_1, \ldots, x_n \rightarrow b'_m)$.

Если предикатные символы в каждой паре одинаковы ($i = j$ и $\alpha = \beta$), то рассуждения аналогичны.

Д о с т а т о ч н о с т ь.

Покажем, что невыполнение любого из перечисленных условий влечёт неизоморфность пар.

То, что при несовпадении предикатных символов или если литералы отличаются отрицаниями пары не изоморфны, очевидно.

Если количество констант с числом вхождений s в одну из пар отличается от количества констант с тем же числом вхождений в другую пару, то построение унификатора невозможно.

Определение. *Индекс пары – это минимальное количество литералов в базе данных, конъюнкция которых изоморфна данной паре.*

Определение. *Адрес пары – это список адресов в базе данных входящих в нее литералов.*

Алгоритм выделения множеств литералов, конъюнкции которых изоморфны

0. $l := 0$.

1. Создадим список всевозможных конъюнкций пар литералов. Для этого попарно перебираем все литералы из базы данных и храним вместе с каждой парой ее адрес (список из двух чисел – номеров литералов в базе данных). Литералы в паре считаем упорядоченными по возрастанию номеров предикатных символов.

2. Для каждых двух непомеченных пар с одними и теми же предикатными символами проверяем их на изоморфность. Для этого сравниваем списки их аргументов.

• Если списки разной длины, то пары не изоморфны.

• Если найдется аргумент в одной из пар, который входит в пару более одного раза, при этом ему соответствуют разные аргументы в другой паре, то пары не изоморфны. Иначе пары изоморфны.

• Если предикатные символы у литералов одинаковые, повторяем процедуру, поменяв литералы одной из пар местами.

3. Найденные изоморфные пары помечаем как проверенные, создаем пару с переменными вместо аргументов, заносим адреса пар с литералами в список адресов пары с переменными, добавляем унификаторы к списку подстановок.

4. Перебираем остальные пары в поисках изоморфных паре с переменными, применяя пункты 2-3.

5. Повторяем пункты 2-4, игнорируя уже проверенные пары.

6. Удаляем проверенные пары, сортируем список по убыванию индекса.

7. Берем пару с максимальным индексом($Pmax$) и вводим новый предикатный символ (l+1)-го уровня (где l – максимальный уровень предикатных символов, входящих в пару), определяемый этой парой. Вместе с символом храним полную форму пары.

8. Ищем пары, имеющие пересечения с $Pmax$. Для этого рассмотрим отдельно каждую пару, имеющую с ней пересекающиеся адреса($Pinti$).

9. Поочередно подставляя в $Pinti$ соответствующие общим адресам унификаторы, смотрим, имеются ли у полученных конъюнкций общие литералы.

10. Если да, то добавляем конъюнкцию этих пар(3 литерала предыдущих уровней) к списку в виде пары с предикатным символом наибольшего из уровней исходных пар + 1, при этом сохраняем полную форму конъюнкции.

11. Выделяем изоморфные пары среди только что добавленных в п.10.

12. Обновляем индексы пар (если с учетом унификаторов пара является подформулой ранее полученного объединения, удаляем адрес из списка ее адресов), удаляем пары с нулевым индексом.

13. Повторяем процедуру с пункта 7, пока остаются пары с индексом 2.

14. Добавляем новые предикатные символы и пары, содержащие их, в базу данных.

Алгоритм реализован на языке C++.

Этот алгоритм имеет экспоненциальную от длины записи базы данных оценку вычислительной сложности, так как количество пар 1-го, 2-го и т.д. уровней растёт экспоненциально. Этот алгоритм применяется однократно.

Заключение

Использование построенной многоуровневой базы данных для решения задачи КБЗ имеет существенно меньшую оценку вычислительной сложности, чем её непосредственное решение, что позволяет уменьшить оценки времени при её многократном решении. Изменение таких оценок при многократном решении задачи такое же как при решении задачи ВКИ при использовании многоуровневой системы описания целевых условий.

ЛИТЕРАТУРА

1. Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. М. : Мир, 1982.
2. Косовская Т.М. Многоуровневые описания классов для уменьшения числа шагов решения задач распознавания образов, описываемых формулами исчисления предикатов // Вестн. С.-Петербург.ун-та. Сер. 10. 2008. Вып. 1. С. 64–72.

3. Косовская Т.М. Частичная выводимость предикатных формул как средство распознавания объектов с неполной информацией // Вестн. С.-Петербург.ун-та. Сер. 10. 2009. Вып. 1. С. 74–84.
4. Косовская Т.М. Некоторые задачи искусственного интеллекта, допускающие формализацию на языке исчисления предикатов, и оценки числа шагов их решения // Труды СПИИРАН, 2010. Вып. 14. С. 58–75.
5. Петров Д.А. Алгоритмы выделения наибольшей общей с точностью до имён переменных формул исчисления предикатов и их реализация // Материалы 9-й конференции «Информационные технологии в управлении (ИТУ – 2016). СПб.: АО «Концерн «ЦНИИ «Электроприбор», 2016. С. 97–102.
6. Коба Д.А. Алгоритм построения многоуровневого описания в задачах распознавания образов и оценка числа шагов его работы // Материалы XVII Международной школы-семинара "Синтез и сложность управляющих систем" им. акад. О.Б. Лупанова (Новосибирск, 27 октября – 1 ноября 2008 г.), Изд-во Института математики, Новосибирск, 2008. С. 55–59.

T.M. Kosovskaya (St. Petersburg State University)

Construction of a level data base for decreasing of Conjunctive Boolean Query computational complexity

Two NP-complete problems CONJUNCTIVE BOOLEAN QUERY (concerning databases) and SATISFIABILITY IN A FINITE INTERPRETATION (concerning solving some Artificial Intelligence problems) which have almost identical formulations but essentially different applications are compared in the paper. Construction of a level database, permitting to decrease computational complexity of the first one is described. This construction is based on the extraction of sets of literals, conjunctions of which in every set are isomorphic to each other, from a database. Algorithms and their computational complexities are presented.

КОГНИТИВНЫЕ ТЕХНОЛОГИИ В УПРАВЛЕНИИ МОНИТОРИНГОМ

В. И. Осипов, Н. А. Жукова
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук, 14 линия 39, тел. 8 (812) 328-34-11, e-mail: osipov_vasiliy@mail.ru, nazhukova@mail.ru*)

М. А. Червонцев, А. И. Водяхо
(*Санкт-Петербургский государственный электротехнический университет, ул. проф. Попова 5, тел. 8 (812) 346-44-87, e-mail: aivodyaho@mail.ru, machervontsev@mail.ru*)

Н. В. Климов
(*ИТМО, Кронверский пр. 49, 8 (812) 232-97-04, e-mail: hocico16@gmail.com*)

Аннотация

Увеличение числа используемых технических объектов, их сложности, объема поступающих от них данных определяют новые требования к системам мониторинга. В докладе рассматривается подход к построению перспективных систем когнитивного мониторинга. Реализация когнитивных возможностей систем обеспечивается за счет машинного синтеза процессов мониторинга. Предлагается формализованная постановка задачи синтеза и алгоритм ее решения. Отдельное внимание уделяется технологиям реализации синтезированных процессов на практике. Приводится развернутый пример синтеза процессов когнитивного мониторинга.

Ключевые слова: системы мониторинга, когнитивные системы, методы синтеза

Введение

Под искусственными когнитивными системами понимаются системы, способные воспринимать и связывать события (сигналы), формировать их пространственно-временные модели и использовать их для выработки и реализации управляющих решений. С учетом этого определения, когнитивность искусственных систем тесно связана с традиционным мониторингом (процессом наблюдения за внешним миром и обработкой результатов наблюдения). Творческие задачи в традиционном мониторинге преимущественно возлагаются на высококвалифицированных экспертов. При их решении востребованными оказываются многие модели и методы искусственного интеллекта. В настоящее время из-за увеличивающегося числа наблюдаемых объектов, их сложности, объема поступающих от них данных, возможностей экспертов для решения задач мониторинга оказывается недостаточно. В связи с этим возникает проблема наделения систем мониторинга машинной когнитивностью. При этом машинный когнитивный мониторинг должен согласованно также решать задачи сбора данных с наблюдаемых объектов и обработки полученных данных в интересах достижения поставленной цели мониторинга. Когнитивное решение этих задач требует перехода от традиционно используемых видов наблюдения и обработки информации к машинному когнитивному мониторингу.

Известные подходы [1-7] к когнитивному мониторингу во многом не проработаны. Когнитивный мониторинг только сейчас по настоящему начинает зарождаться. В интересах развития методов такого мониторинга в статье предлагается новая формализованная постановка задачи автоматического синтеза соответствующих процессов мониторинга, а также алгоритм ее решения и технологии реализации на практике.

Математическая постановка задачи

Исходными данными для синтеза процессов мониторинга является модель M наблюдаемого объекта. Модели объекта описываются в пространстве состояний S этого объекта и связей между ними: $M = \{S, R\}$. Модели представляют собой систему взаимосвязанных моделей, распределенных по нескольким уровням: $\{M^0\}, \{M^1\}, \ldots, \{M^l\}, \ldots, \{M^K\}$, где K - число уровней.

Синтез моделей наблюдаемых объектов осуществляется на основе поступающих от них потоков данных. Информационное содержание потоков носит название контента K. Контент информационного потока свой на каждый момент времени $K(t)$. Объем информации, полученный

об объекте на некоторый момент времени $V(K,t)$, - это случайная величина, которая имеет кумулятивную функцию распределения $F_V(V(t))$.

На синтез моделей объектов влияют условия мониторинга. В разных условиях для одного и того же состояния объекта могут строится разные модели. Условия мониторинга определяют контекст мониторинга $C(t)$.

Тогда модель объекта на некоторый момент времени t_k имеет вид $M(t_k) = \{S(K(t_k), C(t_k)), R(C(t_k), C(t_k))\}$. В зависимости от контента и контекста модели могут отличаться как по структуре, так и по наполнению.

Синтез процесса мониторинга предусматривает, что имеется исходная модель объекта M_s и целевая модель объекта M_w. Требуется построить процесс мониторинга $P: M_s \to M_t$. Наличие нескольких уровней в моделях объектов требует синтеза многоуровневых процессов.

При построении процесса мониторинга могут учитываться временные ограничения. В этом случае определяется временной диапазон $t_k \in [t_s; t_w]$. С учетом этого процесс $P: M(t_s) \to M(t_k)$.

Синтезируемый процесс оценивается в терминах сокращения информационного разрыва между исходной и целевой моделями [8]. Таким образом, суть процесса мониторинга состоит в том, чтобы преобразовать исходную модель M_s таким образом, чтобы объем информации I в модели M_s соответствовал объему информации модели M_w: $I(M_s) \equiv I(M_w)$

Для синтеза процессов мониторинга необходимо определить входные и выходные данные, условия, их связывающие.

Входные и выходные данные при многоуровневом синтезе разделены на группы по принятым уровням иерархии описания процессов. Входные данные задаются как $\{d_{s0}\}, \{d_{s1}\}, \ldots, \{d_{sl}\}, \ldots, \{d_{sK}\}$, выходные - $\{d_{w0}\}, \{d_{w1}\}, \ldots, \{d_{wl}\}, \ldots, \{d_{wK}\}$.

Условия синтеза определяются следующим образом [7]:

$$F_{zv}^i(d_{zv_e}^i; e = \overline{1, E_z}) \to d_{zv_a}^i; z = \overline{1, Z_i};\ v = \overline{1, V}_{iz};\ i = \overline{1, K}; \tag{1}$$

$$F_{zv}^i(\cdot) \in FB^i; \tag{2}$$

$$d_{sl}, d_{zv_e}^l, d_{zv_a}^l, d_{wl} \in DB^l;\ \boldsymbol{i = \overline{1, K}}, \tag{3}$$

В (1) – (3) приняты обозначения: $F_{zv}^i(\cdot)$ - функции z-х типов и v-х видов, которые могут реализовываться процессом на i-м уровне иерархии; $d_{zv_e}^i, d_{zv_e}^i$ - данные; K - число уровней иерархии; FB^l, DB^l - множество допустимых функций и множество допустимых данных на i-м уровне.

Согласно (1) на i-м уровне иерархии с использованием функции $F_{zv}^i(\cdot)$ при известных $d_{zv_e}^l$ могут быть определены данные $d_{zv_\epsilon}^l$. Выражения (2), (3) отражают ограниченность возможных функций и данных на i-м уровне.

С учетом этого искомый процесс PRG представляется в виде:

$$PRG = \left\{F_{zv}(d_{zv_e}; e = \overline{1, E_z}) \to d_{zv_a};\ \{d_s\};\ \{d_w\};\ z = \overline{1, Z};\ v = \overline{1, V_z}\right\}. \tag{4}$$

Если поставить всем элементам в условиях (2) соответствующие им предикаты, $P_{zv0}(F_{zv0}(\square))$, $P_{zv1}(d_{zv1})$, …, $P_{zva}(d_{zva})$, принимающие значение 1, когда переменные определены (истины), и 0 в противном случае, то условия задачи могут быть записаны в виде $\{P_{zv0} \wedge P_{zv1} \wedge, ..., P_{ZVE_z} \to P_{zva} \mid S_{zv}, M_{zv}, z = \overline{1, Z} : v = \overline{1, V}\}$, где S_{zv} - статус zv-го условия; M_{zv} - множество номеров основных условий, при которых предусловия и постусловия, соответственно, истинны и ложны. Для основных условий $M_{zv} = \varnothing$. С учетом этого основное условие со свободными переменными выполнимо, если $P_{zv0} \wedge P_{zv1} \wedge, ..., P_{ZVE_z} = 1$. Вспомогательное условие разрешимо, когда и правая часть в нем истина, $P_{zv0} \wedge P_{zv1} \wedge, ..., P_{ZVE_z} \wedge P_{zva} = 1$.

При синтезе моделей процессов исходные и результирующие данные, а также функции и условия их выполнения задаются исходя их моделей объектов. При этом строится следующая система отображений

$$\begin{Bmatrix} M^i(\overline{d}_r^i) \xrightarrow{F_i^{DB}} DB^i(\overline{d}_r^i) \\ M^i(\overline{d}_r^i) \xrightarrow{F_i^{FB}} FB^i(\overline{d}_r^i) \\ M^i(\overline{d}_r^i) \xrightarrow{F_i^{P}} P^i(\overline{d}_r^i) \end{Bmatrix} \quad (5)$$

В (5) рассмотрены отображения для i-го уровня. Функция F_i^{DB} позволяет строить пространство данных для синтеза процессов, F_i^{FB} и F_i^{P} - обеспечивает формирование множества условий и соответствующих им предикатов. Отображения рассматриваются для r-го шага синтеза процесса.

При выборе очередного шага мониторинга оценивается прирост информации $\Delta I(M(d_r),M(d_{r+1}))$, имеющейся на шаге r модели, который может быть получен за счет очередного шага синтеза. Задача выбора способа синтеза может быть сформулирована следующим образом:

$$V_r^{opt}(M(d_r)) = \max_{r\in\Omega} \Delta I(M(d_r),M(d_{r+1})). \quad (6)$$

В результате выполнения очередного шага мониторинга могут быть получены новые данные или выявлены новые связи, установлены новые условия, при которых связи проявляются. Эти результаты используются для уточнения модели объекта. При этом на i-ом уровне выполняется:

$$\begin{Bmatrix} DB^i(\overline{d}_{r+1}^i) \xrightarrow{\tilde{F}_i^{DB}} M^i(S)(\overline{d}_{r+1}^i) \\ FB^i(\overline{d}_{r+1}^i) \xrightarrow{\tilde{F}_i^{FB}} M^i(R)(\overline{d}_{r+1}^i) \\ P^i(\overline{d}_{r+1}^i) \xrightarrow{\tilde{F}_i^{P}} M^i(\overline{d}_{r+1}^i) \end{Bmatrix} \quad (7)$$

где $\tilde{F}_i^{DB}$, $\tilde{F}_i^{FB}$, $\tilde{F}_i^{P}$ - функции, обеспечивающие расширение и наполнение модели в соответствии с результатами выполненного шага мониторинга

В следующем разделе предлагается алгоритм, обеспечивающий синтез процессов когнитивного мониторинга.

Алгоритм решения

Входные и выходные данные алгоритма определяются в соответствии с множествами $\{d_{s0}\},\{d_{s1}\},\ldots,\{d_{sl}\},\ldots,\{d_{sK}\}$ и $\{d_{w0}\},\{d_{w1}\},\ldots,\{d_{wl}\},\ldots,\{d_{wK}\}$. Данные представляются в виде вектора фактов, например: [a1, a2, …, ak, c1, c2, … ck]. Входные данные определяют исходные факты, выходные - искомые.

Функции переходов F (*function*) из одного состояния в другое задаются в виде следующей структуры: $[conditions]F([args]) \rightarrow [result]$, где $conditions$ - условия, определяющие область применения функции, $args$ - вектор, имеющихся на момент выполнения функции фактов, характеризующее начальное состояние, $result$ - результат работы функции, включающий один или несколько новых фактов. Исходные и новые факты определяют результирующее состояние. Результирующее состояние оказывается целевым, если доказаны все искомые факты.

Условия переходов задаются в виде логических выражений и могут быть как статическими, так и динамическими. Статические условия задаются до начала синтеза, динамические могут определяться на каждом из шагов синтеза.

Уровни моделирования задаются в виде иерархической графовой структуры. В качестве такой структуры выступает дерево или лес. Структура представляет собой вектор элементов, каждый из которых имеет вид: $[F, level, number, parent]$, где F - это набор функций перехода,

заданные для уровня, на котором выполняется синтез, $level$ - номер уровня, $number$ - порядковый номер элемента на уровне, $parent$ - связь с родительским элементом, который находится на вышестоящем уровне. При этом факты, доказанные на дочерних уровнях, могут быть использованы в качестве исходных фактов для функций переходов на родительских уровнях.

Таким образом, задачи синтеза сводится к нахождению такой последовательности функций переходов F , при котором искомые факты будут доказаны..

Предлагаемый алгоритм предусматривает реализацию двунаправленной многоуровневой процедуры формирования последовательности функций F , обеспечивающих доказательство искомых фактов исходя из исходных. Это соответствует поиску переходов (пути) из начального состояния процесса в целевое. На первом шаге выполняется построение обратного пути, т.е. доказательство исходных фактов в предположении, что заданы искомые факты. Факты, определенные для функций переходов, при включении их функций в путь, доказываются рекурсивно. На втором шаге осуществляется восстановление прямого пути на основе полученного обратного пути. Уровни процедуры определяются уровнями моделей и процессов мониторинга. Поиск пути выполняется начиная с верхнего уровня. Переход на более низкие уровни осуществляется в том случае, если на текущем уровне один или несколько фактов не могут быть доказаны. Более подробное описание основных шагов алгоритма приведено ниже.

Алгоритмом предусматривается выполнение следующих шагов:

Шаг 1. Выполняется поиск пути для доказательства целевого факта (для достижения целевого состояния процесса) на нулевом уровне моделирования. Для этого:

Шаг 1.1. Среди множества исходных фактов уровня осуществляется поиск целевого факта. Если такой факт найден, то алгоритм завершает работу.

Шаг 1.2. Если целевой факт не найден, то среди всех функций переходов выбирается такая функция, вектор выходных данных которой содержит целевой факт;

Шаг 1.3. Если функция найдена, выполняется проверка выполнимости условий ее применения. При проверки могут использоваться доказанные факты, соответствующие текущему состоянию процесса;

Шаг 1.4. Если условия выполняются, то в последовательность функций переходов, сформированную к текущему шагу (в текущий путь) добавляется найденная функция. Для каждого входного факта выполняется аналогичный рекурсивный поиск пути, начиная с Шага 1.

Шаг 1.5. Если условия не выполняются, либо функция не найдена, аналогично производится рекурсивный поиск пути к целевому состоянию на всех дочерних уровнях модели.

Шаг 1.6. Если путь к целевому состоянию на дочерних уровнях найден, то такой путь прибавляется к текущему пути. Если не найден, то поиск завершается. Результатом такого поиска является пустой путь.

Шаг 2. В случае, если найден непустой путь к целевому состоянию, то восстанавливается прямой путь от начального состояния, определяемого множеством исходных фактов, к целевому. Для этого реверсивно меняется порядок функций в найденном пути. Дублирующие функции исключаются.

Путь, построенный в результате выполнения такой процедуры, может использоваться для формирования управляющих воздействий на объекты.

Описанный алгоритм можно представить в виде следующего псевдокода:

```
1. Find path to target condition T on item I
2. path is empty
3. For each i in input data
4.  if (i == T)
5.   path is not feasible
6.  else
7.   For each F in functions
8.      if (F[output] == T and F[conditions] are satisfied)
9.         For each fi in F input data
10.           iPath = Find path to target condition fi on item I
```

```
11.          if (iPath is null)
12.            For each ci in child items
13.              iPath = Find path to target condition fi on item ci
14.          if (iPath is null)
15.            path is not feasible
16.          else
17.            add iPath to path
18.    return path
19.
20. p = Find path to target condition T on root item I0
21. reverse p order
22. remove duplicated functions in p
23. result is optional path to condition T
```

Технологии синтеза процессов мониторинга

Когнитивные технологии – это совокупность моделей, методов и средств, а также способов их применения, которые обеспечивают реализацию моделей процессов когнитивного мониторинга в информационных системах. При этом строится система переходов, в результате которых формальные абстрактные описания процессов представляются в виде программных скриптов, которые могут быть выполнены средствами информационных систем.

Система переходов предполагает раскрытие таких абстрактных понятий как модели объектов и их состояния, а также способов преобразования моделей, обеспечивающих переходы от исходных моделей к целевым. Можно выделить две основные группы переходов. К первой группе относятся переходы, которые позволяют представить процессы, описанные на языке моделирования, в виде стандартизованных моделей бизнес процессов. Вторая группа переходов обеспечивает генерацию исполняемых процессов в соответствии с бизнес моделями.

Технология построения программно реализуемых процессов предусматривает наличие следующей информации.

1. Формализованное описание предметной области. Такое описание должно содержать информацию об объектах, их особенностях, решаемых задачах, условиях их решения и иную, которая может учитываться при мониторинге.

2. Формализованное описание моделей, методов и средств представления и обработки данных. Описание представляет собой систему таксономий, каждый элемент которой сопровождается спецификацией, интерпретируемой на машинном уровне. Так, для методов могут указываться входные и выходные данные, пред- и пост- условия, область применения, критерии оценки и иные.

3. Логические правила, определяющие допустимые способы обработки данных, поступающих от объектов. При этом устанавливаются связи между элементами формализованных описаний предметной области и области обработки данных Правила учитывают состав и свойства данных, получаемых от объектов и их объем.

Тогда при синтезе программных моделей процессов построение пространства данных и пространства функций процессов мониторинга, предусмотренное в (5) сводится к получению из формализованной модели предметной области сущностей, описывающих элементы объекта и отношения, установленные между этими элементами.

При построении переходов между состояниями синтезированных процессов предусматривается выбор методов и алгоритмов, которые обеспечивают преобразование моделей, детализация синтезируемых переходов.

При построении переходов между состояниями синтезированных процессов предусматривается детализация способов преобразования моделей объектов, предусмотренных этими переходами. При этом осуществляется выбор конкретных методов и алгоритмов обработки данных и их настройка. Для выбора методов используются логические правила. Настройка методов осуществляется исходя из спецификации метода и входных данных.

Оценка переходов и поиск наилучшего в соответствии с (6) основана на оценке информативности полученных результатов обработки. Оценка проводится с использованием критериев информативности, определенных для методов.

Учет полученных результатов в моделях объектов (7) предполагает включение этих данных в описания объектов предметной области. При этом может уточняться структура моделей, добавляться новая информация о свойствах моделей и т.д.

Ниже пример синтеза процессов мониторинга ракетно-космического носителя ‘СОЮЗ-2’ при подготовке к проведению испытаний.

Пример синтеза процессов мониторинга

Рассмотрим пример синтеза процессов мониторинга при подготовке объектов ракетно-космической техники (РКТ) к проведению испытаний.

В настоящее время подготовка объектов к проведению испытаний представляет собой серию мероприятий в рамках которых проводятся многие сеансы измерений параметров испытуемых объектов. В зависимости от цели мониторинга определяется состав измеряемых параметров и продолжительность сеансов измерений. При проведении измерений на объект могут оказываться управляющие воздействия в виде входных команд. Передача команд может производиться по многим каналам. Команды могут выдаваться на протяжении некоторого интервала времени или быть импульсными. Предусматривается возможность проведения сеансов измерений в различных режимах. Режимы определяют уровни выдаваемых входных сигналов. Результаты измерений, получаемые при проведении сеансов, используются для оценки состояния исследуемого объекта и его элементов. На основе результатов обработки данных, полученных в рамках отдельных сеансов, могут корректироваться состав сеансов, планируемых к проведению в дальнейшем или формироваться новые сеансы.

Для формализованного описания объектов РКТ используются онтологические модели. В этих моделях содержится как информация об элементах этих объектов, так и о взаимосвязях между элементами. Для обработки данных, поступающих об объектах, используются свои методики для каждого из типов объектов. Основным источником информации об объектах является конструкторская документация, представляемая организациями, занимающимися проектированием и разработкой объектов, а также экспертные знания специалистов, ответственных за их эксплуатацию.

При синтезе моделей процессов мониторинга предлагается реализовывать следующую логику. Из множества параметров, которые могут быть измерены, выделяются группы параметров, позволяющие оценить состояние каждой из систем исследуемого объекта. Процессы, обеспечивающие сбор и обработку таких параметров, будем относить к процессам нулевого уровня. При выявлении отклонений в поведении одного из параметров, определяется принадлежность этого параметра к структурному элементу объекта. Синтезируются процессы сбора и обработки значений параметров, измеряемых на этом элементе. Такие процессы – это процессы первого уровня. Определив в структурном элементе агрегаты и узлы, в работе которых наблюдаются отклонения, могут проводиться сеансы измерений параметров этих узлов и агрегатов в различных режимах. Эти процессы определяют процессы второго уровня.

Логика синтеза формируется в соответствии с описанием предметной области. В описании определяются связи между объектами и их элементами, которые составляют основу для синтеза процессов. За счет добавления новых связей или изменении существующих могут перестраиваться реализуемые процессы мониторинга или синтезироваться новые.

Для рассмотренного способа синтеза процессов должны быть установлены связи между измеряемыми параметрами исследуемого объекта, его системами и конструктивными элементами.

Приведем примеры процессов, синтезируемых для мониторинга РКН ‘СОЮЗ-2’. В конструкции РКН выделяется пять блоков – один центральный блок (блок А) и четыре боковых блока (блоки Б-Д). ‘СОЮЗ-2’ включает в состав несколько систем, в том числе, систему подачи топлива. Элементы этой системы размещаются на блоках А-Д. Для оценки состояния системы подачи топлива контролируются параметры, часть из которых приведена в таблице 1.

Таблица 1

Короткое имя	Имя	Обозначение
ДБО	Давление в баке окислителя	$P_{ДБО}$
ДБГ	Давление в баке горючего	$P_{ДБГ}$
ДБА	Давление в баке азота	$P_{ДБА}$
ДБП	Давление в баке пероксида водорода	$P_{ДБП}$

При выявлении отклонений в поведении какого-либо из параметров на одном из блоков, проводится сбор и обработка других параметров, отражающих состояние именно этого блока. При этом могут рассматриваться как параметры, относящиеся к возможно неисправному элементу, так и к другим элементам этого блока. В таблице 2 приведены примеры таких параметров, для случая, когда выявлены отклонения в значениях параметра давления в баке пероксида водорода в блоке Б.

Таблица 2

Короткое имя	Имя	Обозначение
ТКРПб	Температура корпуса редуктора пероксида водорода в блоке Б	$P_{ТКРПб}$
ТБПб	Температура в баке пероксида водорода в блоке Б	$P_{ТБПб}$
ТМБОб	Температура среды в межбаковом отсеке в блоке Б	$P_{ТМБОб}$
ДАДб	Давление азота в демпферах	$P_{ДАДб}$
ДВБб	Давление в воздушных баллонах	$P_{ДВБб}$
ВПРб	Давление воздуха после редуктора точной настройки	$P_{ВПРб}$

В случае, если по результатам обработки значений параметров выявлено, что отклонения также наблюдаются в поведении параметра температуры в баке пероксида водорода, то может быть определен перечень дополнительных режимов проверки для уточнения причин возникновения отклонений в значениях параметров. Описание одного из режимов приведено в таблице 3.

Таблица 3

Режим	Команда	Канал	Вид
Режим 1	Вкл. Ф1ТПС2	0,6А №4	Выдать
	Выкл. Ф1ТПС2	0,6А №4	Снять

В результате синтеза процессов мониторинга был построен трехуровневый процесс. Множество состояний процесса приведены в таблице 3, множество функций переходов – таблице 4. Синтезированный процесс представлен на рис. 1. В таблице 4 используются следующие обозначения: F - функция переходов; g - функция обработки результатов измерений параметров. $g \in \{0;1\}$, $g(\mathrm{P}) \equiv 0$, если в значениях параметров отклонений не выявлено, $g(\mathrm{P}) \equiv 1$ - в иных случаях; $\overline{g}(\mathrm{P})$ используется для обозначения $g(\mathrm{P}) \equiv 1$.

Таблица 3

Уровень	Обозначение	Описание
0	s00	начальное состояние
0	w00	конечное состояние; отклонений состоянии объекта по результатам мониторинга не выявлено
0	w01	конечное состояние; выявлены отклонения в состоянии объекта по результатам мониторинга; выполнен сбор данных для выявления причин и устранения отклонений
0	c00	получение параметров ДБО, ДБГ, ДБА, ДБП
0	c01	обработка параметров ДБО, ДБГ, ДБА, ДБП
0	c02	обработка параметров системы подачи топлива прошла успешно; в поведении параметров ДБО, ДБГ, ДБА, ДБП отклонений не выявлено
0	c03	выявлены отклонения в поведении параметра ДБП системы подачи топлива
1	c10	получение параметров ТКРПб, ТБПб, ТМБОб, ДАДб, ДВБб, ВПРб
1	c11	обработка параметров ТКРПб, ТБПб, ТМБОб, ДАДб, ДВБб, ВПР
1	c12	обработка параметров блока Б прошла успешно; в поведении параметров ТКРПб, ТБПб, ТМБОб, ДАДб, ДВБб, ВПРб отклонений не выявлено
1	c13	выявлены отклонения в поведении параметра ТБП блока Б
2	c20	получение параметров ДБПб и ТБПб в режиме ‘Режим 1’

Таблица 4

Условие	Функция	Описание
-	$F(s00) \to c00$	сбор данных
-	$F(c00) \to c01$	обработка данных
$g(P_{ДБО}) \wedge g(P_{ДБГ}) \wedge g(P_{ДБА}) \wedge g(P_{ДБП})$	$F(c01) \to c02$	анализ результатов обработки данных
$g(P_{ДБО}) \wedge g(P_{ДБГ}) \wedge g(P_{ДБА}) \wedge \overline{g}(P_{ДБП})$	$F(c01) \to c03$	
-	$F(c03) \to c10$	сбор данных
-	$F(c10) \to c11$	обработка данных
$g(P_{ТКРПб}) \wedge g(P_{ТБПб}) \wedge g(P_{ТМБОб}) \wedge g(P_{ДАДб}) \wedge \wedge g(P_{ДВБб}) \wedge g(P_{ВПРб})$	$F(c11) \to c12$	анализ результатов обработки данных
$g(P_{ТКРПб}) \wedge \overline{g}(P_{ТБПб}) \wedge g(P_{ТМБОб}) \wedge g(P_{ДАДб}) \wedge \wedge g(P_{ДВБб}) \wedge g(P_{ВПРб})$	$F(c11) \to c13$	
-	$F(c13) \to c20$	сбор данных
-	$F(c20) \to w01$	завершение процесса; формирование выходных данных
	$F(c02) \to w00$	
	$F(c12) \to w00$	

В результате выполнения процесса мониторинга в рассматриваемых условиях выходные данные включают следующую информацию о наличии отклонений в состоянии системы подачи топлива, а также значения параметров $P_{ДБП}$ и $P_{ТБПб}$ блока Б, измеренные в режиме 1 ‘Режим 1’.

Описание функций системы на языке верхнего уровня:

```
[
 {
    "id": "baad85ac-1733-4f25-9609-e335807bbb4c", "parent": null,
    "level": 1,  "number": 1, "functions": [
      { "args": ["c20"], "result": "w01","conditions": "" },
      { "args": ["c21"],  "result": "w00","conditions": "" }
    ]
  },
  {
    "id": "01dc2d50-d3b5-4710-bb7b-3dee2048f8ce", "parent": "baad85ac-1733-4f25-9609-
e335807bbb4c",
    "level": 2, "number": 1, "functions": [
      { "args": [ "c13" ], "result": "c20", "conditions": "" },
      { "args": [ "c12" ], "result": "c21", "conditions": "" },
      { "args": [ "c11" ], "result": "c13", "conditions": "
```

$$g(P_{ТКРПб}) \wedge \overline{g}(P_{ТБПб}) \wedge g(P_{ТМБОб}) \wedge g(P_{ДАДб}) \wedge \wedge g(P_{ДВБб}) \wedge g(P_{ВПРб})$$

```
" },
      { "args": [ "c11" ], "result": "c12", "conditions": "
```

$$g(P_{ТКРПб}) \wedge g(P_{ТБПб}) \wedge g(P_{ТМБОб}) \wedge g(P_{ДАДб}) \wedge \wedge g(P_{ДВБб}) \wedge g(P_{ВПРб})$$

```
" }
    ]
  },
  {
    "id": "ecef1866-c68a-468b-bcfb-e1fc6c4a55b9", "parent": "01dc2d50-d3b5-4710-bb7b-
3dee2048f8ce",
    "level": 3, "number": 1, "functions": [
      { "args": [ "c03" ], "result": "c10", "conditions": "" },
      { "args": [ "c01" ], "result": "c03", "conditions": "" },
      { "args": [ "c01" ], "result": "c02", "conditions": "" },
      { "args": [ "c00" ], "result": "c01", "conditions": "" },
      { "args": [ "s00" ], "result": "c00", "conditions": "" }
    ]
  }
]
```

Предположим, что возникла неисправность в системе подачи топлива в результате которой наблюдаются отклонения в поведении параметров ДБПб и ТБПб.

Попробуем доказать, что при данных условиях система имеет нормальное состояние.

Целевой факт - w00 (нормальное состояние). Он достижим только при возможности достижения фактов c12 или c02. c12 достижимо при наличии c11 и условии, что значения параметра ТБПб не содержат отклонений, c02 - при наличии c01 и отсутствии отклонений в значениях ДБПб. Других возможных путей к w00 не существует, так что нормальное состояние системы не может быть доказано (недостижимо) при наличии указанных условий.

Приведем полный путь доказательства неисправности системы (пути перехода к состоянию w01): $F(s00) \to c00$; $F(c00) \to c01$; $F(c01) \to c03$; $F(c03) \to c10$; $F(c10) \to c11$; $F(c11) \to c13$; $F(c13) \to c20$; $F(c20) \to w01$.

Заключение

В докладе предлагается новая технология, обеспечивающая возможность управления процессами мониторинга на машинном уровне. В основу технологии положены модели и методы многоуровневого синтеза. Разработан и представлен программный алгоритм многоуровневого синтеза, который обеспечивает использование абстрактного аппарата синтеза при решении практических задач. Применение этого алгоритма совместно с существующими готовыми решениями, предоставляемыми областью IT, обеспечивает синтез абстрактных формальных моделей процессов мониторинга, формирование бизнес моделей этих процессов с возможностью последующей генерации программного кода.

Перспективным направлением развития технологий когнитивного мониторинга является переход к активному использованию использование в качестве опорных технологии Process Mining и технологии Model Mining. Первая технология может позволить строить новые модели, описывающие объекты, их структуру и свойства, вторая – формализовывать процессы, происходящие на объектах, выявлять их особенности. Построение таких структурных и поведенческих моделей может обеспечивается на машинном уровне. Управление построением и перестройкой моделей основано на использовании механизмов адаптации. Информационные системы мониторинга, построенные на основе машинных моделей могут эффективно использоваться как системы поддержки принятия решений по управлению объектами.

ЛИТЕРАТУРА

1. Augello A., Infantino I., Pilato G., Vella F. Creativity evaluation in a cognitive architecture // Biologically Inspired Cognitive Architectures. V. 11, 2015 Pp. 29-37
2. Lieto A., Cruciani M. Introduction to cognitive artificial systems. Connection Science. V.27(2), ·2015
3. Kotseruba I., Tsotsos J. Review of 40 Years of Cognitive Architecture Research: Core Cognitive Abilities and Practical Applications. ArXiv:1610.08602 [cs.AI], 2018
4. Sun R. The importance of cognitive architectures: an analysis based on CLARION. Journal of Experimental & Theoretical Artificial Intelligence. V. 19(2), 2007
5. Glodek M., Honold F., Geier T., Krell G., Nothdurft F., Reuter S., Schüssel F., Hörnle T., Dietmayer K., Minker W., Biundo S., Weber M., Palm G., Schwenker F. Fusion paradigms in cognitive technical systems for human–computer interaction. Neurocomputing .V.161, 2015
6. Goertzel B., Lian R., Arel I., Garis H., Chen S. World survey of artificial brains. Part II: Biologically inspired cognitive architectures. Neurocomputing. V. 74 (1-3), 2010
7. Haykin S., Fatemi M.; Setoodeh P.; Xue Y. Cognitive Control. Proceedings of the IEEE. V. 100 (12), 2012
8. Haykin S., Xue Y., Setoodeh P. Cognitive Radar: Step Toward Bridging the Gap Between Neuroscience and Engineering. Proceedings of the IEEE . V. 100 (11), 2012

V.I. Osipov, N. A. Zhukova (St. Petersburg Institute of Informatics and Automation of the Russian Academy of Sciences), **M.A, Chervontsev, A.I. Vodyaho** (St. Petersburg State Electrotechnical University), **N.V. Klimov** (ITMO University)

Cognitive technologies in monitoring

The increase in the number of technical objects used, their complexity, and the volume of data coming from them determine new requirements for monitoring systems. The report examines the approach to constructing prospective cognitive monitoring systems. Implementation of the cognitive capabilities of the systems is ensured by the machine synthesis of the monitoring processes. A formal formulation of the synthesis problem and an algorithm for its solution are proposed. Special attention is paid to the technologies of realization of synthesized processes in practice. A detailed example of synthesis of cognitive monitoring processes is given.

МОДЕЛИ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПО УПРАВЛЕНИЮ ДВИЖЕНИЕМ КОСМИЧЕСКОГО АППАРАТА ПРИ ВОЗНИКНОВЕНИИ НЕШТАТНЫХ СИТУАЦИЙ

Е. Н. Алёшин[1], **В. А. Терещук**[2], **С. А. Терещук**[2]
(*Военно-космическая академия имени А. Ф. Можайского,
С.-Петербург, ул. Ждановская, д. 13, aleshin_evgeny@inbox.ru*)

Аннотация

Особую важность при эксплуатации космических аппаратов приобретает задача обеспечения усиления поддержки процессов выработки управленческих решений на основе комплексной автоматизации для повышения эффективности его функционирования в условиях возникновения нештатных ситуаций.

В данной статье рассмотрены вопросы разработки и применения интеллектуальной системы поддержки принятия решений по управлению движением космического аппарата при возникновении нештатных ситуаций. Предлагается место интегрированной системы поддержки принятия решения в автоматизированной системе управления космическими аппаратами, состав экспертной системы поддержки принятия решений и назначение каждого из её элементов, а также технология формирования базы знаний, содержащей информацию о процессе (логике) функционирования системы управления движением космического аппарата.

На основе принципов системного подхода и современных информационных технологий к построению базы знаний экспертной системы поддержки принятия решений описан состав моделей и алгоритмов управления движением космического аппарата в нештатных ситуациях.

Для представления знаний экспертов на различных уровнях формализации (моделирования) в качестве моделей и алгоритмов, позволяющих реализовать различные процедуры вывода, используется семантическая сеть и продукционная система логики функционирования системы управления движением космического аппарата в нештатных ситуациях.

Ключевые слова: система управления движением; космический аппарат; нештатная ситуация.

Введение

Применение традиционных математических моделей и методов для описания процесса функционирования и управления космическими аппаратами (КА) в условиях нештатных ситуаций весьма затруднительно либо невозможно. В связи с этим на первый план выдвигается категория знаний специалистов по управлению, интуиция и опыт которых иной раз являются единственным источником успешности их действий. Процесс выявления экспертных знаний и их формализация явились предпосылкой развития новых информационных технологий (НИТ) в области управления космическими аппаратами как неклассических объектов управления, основными чертами которых являются:

– цель управления такого рода объектами не всегда могут быть точно сформулированы;

– структура (параметр, который мешает автоматизации) таких объектов динамично изменяется;

– процессы управления такими объектами, как правило, являются многошаговыми, и содержание последующих этапов зависит от предыдущих и не всегда может быть спрогнозировано;

– зависимость между параметрами описываемых состояний объектов управления не всегда подается формальному описанию и количественному определению;

– элементы объектов управления могут иметь активную природу и зачастую противодействовать управляемым воздействиям, что особенно характерно для человеко-машинных и организационно-технических систем, а также систем искусственного интеллекта.

Сутью НИТ является удаление различного рода посредников и организация дружелюбного взаимодействия между пользователем и электронно-вычислительной машиной при решении различных задач.

Основными фундаментальными элементами НИТ являются: инженерия знаний, как важнейшее направление искусственного интеллекта (ИИ); развитие средств автоматизации управления, включая вычислительную технику; развитые сетевые технологии.

[1] Кандидат технических наук, старший преподаватель кафедры автоматизированных систем управления космических комплексов.
[2] Курсант ВКА имени А. Ф. Можайского.

Для процессов управления космического аппарата характерно его деление (управления) на четыре вида (контура): управления бортовым ресурсом; управления взаимодействием; управления бортовой аппаратурой; управления движением.

При этом все виды управления взаимосвязаны и обусловлены текущим состоянием объекта. Особо стоит отметить тесную взаимосвязь управления движением, управления бортовой аппаратурой и бортовым ресурсом в рамках системы управления движением космического аппарата.

Анализ нештатных ситуаций возникающих в процессе функционирования КА позволяет выделить три основные группы причин (факторов) их возникновения: космический, антропогенный, техногенный [1].

К группе техногенных факторов относят все происшествия связанные с нарушениями или отказами узлов и элементов бортовых систем КА.

Космический фактор связан с воздействием потенциально опасных воздействий космического пространства.

Антропогенные факторы возникают в результате воздействия человека (оператора) на объект управления.

Вышеперечисленные факторы возникновения нештатных ситуаций, особенности процесса принятия решений по управлению КА на основе НИТ не оставляет сомнений в актуальности данной работы, связанной с разработкой моделей интеллектуальной системы поддержки принятия решений по управлению КА при возникновении нештатных ситуаций.

Методика построения моделей экспертной системы поддержки принятия решений по управлению движением КА при возникновении нештатных ситуаций

Выбор средства автоматизации принятия решений определяется, в первую очередь, особенностями предметной области и характером решаемых в ней задач. Перспективным является использование интегрированных систем поддержки принятия решений (ИСППР), которые включают различные компоненты и используют их в той мере, в которой это диктуется решаемой задачей [2].

ИСППР позволяют оператору дежурной смены оперативно реагировать на нарушения в функционировании КА и принимать меры для парирования нештатных ситуаций (НШС) и восстановление нормального режима работы систем КА. На рис. 1 показано место ИСППР в автоматизированной системе управления КА.

Главным компонентом ИСППР является экспертная система поддержки принятия решений (ЭСППР), которая во взаимодействии с расчетно-логической, интеллектуальной информационно – поисковой, и имитационной системами осуществляет поддержку принятия решений по управлению КА при возникновении НШС.

В состав ЭСППР входят следующие основные функциональные элементы [3], представленные на рис. 2.

1. База данных (БД) о функционировании системы управления движением (СУД) КА в условиях возникновения НШС – содержит все известные режимы функционирования СУД КА.
2. База данных (БД) операций для выхода из НШС – состоит из набора операций, которые необходимо осуществить для парирования НШС.
3. База знаний (БЗ) логики работы СУД КА в условиях возникновения НШС – содержит логические правила вывода [4].
4. Подсистема диалога (ПД), обеспечивающая взаимодействие с ЭСППР на всех этапах ее жизненного цикла и во всех режимах ее функционирования всех категорий пользователей: экспертов (специалистов в предметной области деятельности оператора), инженеров по знаниям (программистов), операторов дежурных смен – конечных пользователей.
5. Подсистема приобретения знаний (ППЗ), позволяющая вносить в БД новые сведения о функционировании СУД.
6. Подсистема объяснений (ПО), разъясняющая оператору, на основе каких рассуждений система получила решение.
7. Решатель, обеспечивающую автоматический вывод решения формируемых пользователем или экспертом задач на основе знаний, хранящихся в базе знаний.



Рис. 1. Интегрированная система поддержки принятия решений и ее место в АСУ КА

Центральной подсистемой ЭСППР, ее системообразующей компонентой является база знаний (БЗ). Предлагается технология построения БЗ, ориентированная на обеспечение ее полноты и непротиворечивости и основанная на разработанной методике многоэтапной (многоуровневой) формализации знаний о процессе функционирования СУД КА. Суть указанной методики заключается в последовательном построении моделей знаний различной степени формализации, начиная с собственно знаний (как модели нулевого уровня) и заканчивая их программной моделью, представляющей собой информационное содержимое БЗ. Такой подход, реализующий принципы системности и мультимодельности при создании БЗ, распространяется на процесс построения других основных функциональных подсистем ЭСППР, что позволяет сформировать соответствующую ей систему моделей и алгоритмов.

Предложены следующие уровни формализации (моделирования), связанные с соответствующими этапами разработки модельно-алгоритмического комплекса ЭСППР: содержательный, структурно-содержательный, структурно-формальный, формальный и программный. В свою очередь, условно содержательный и структурно-содержательный уровни могут быть объединены в обобщенный концептуальный, структурно-содержательный и структурно-формальный – в обобщенный структурный, а структурно-формальный, формальный и программный – в обобщенный формальный уровни моделирования. Перечисленным уровням формализации соответствуют следующие модели подсистем ЭСППР: содержательная модель (СМ), структурно-содержательная модель (ССМ), структурно-формальная модель (СФМ), формальная модель (ФМ) и программная модель (ПМ) [1].

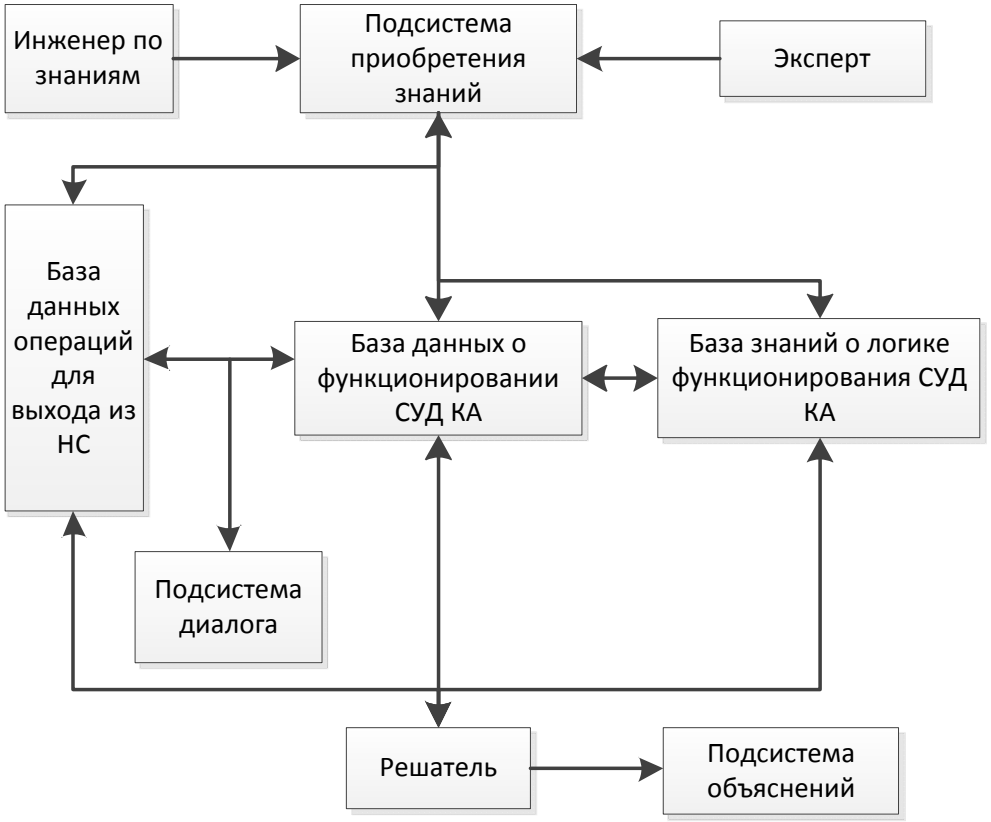


Рис. 2. Структурная схема ЭСППР по управлению движением КА при возникновении нештатных ситуаций

К примеру, СМ БЗ представляет собой вербальное (текстовое) описание знаний.

ССМ БЗ содержит структуру знаний в явном виде, задаваемую соответствующей семантической сетью (СС) (семантической гиперсетью (СГС)).

В качестве СФМ БЗ предлагается использовать соответствующую вложенную сеть Петри (ВСП) [5].

ФМ БЗ представляет собой продукционную модель представления знаний (ПМПЗ).

Переход от ФМ к ПМ БЗ состоит в дальнейшей формализации ПМПЗ в терминах того или иного языка ИЗ, инструментальной среды или оболочки ИС, выбранных для реализации БЗ на программном уровне [6].

Состав системы моделей и алгоритмов ЭСППР, имеющей матричную структуру, представлен табл. 1. Каждый из элементов указанного модельно-алгоритмического комплекса классифицируется по двум признакам: уровню формализации и виду модели (соответствию той или иной подсистеме ЭСППР). При этом БЗ соответствуют модели декларативного (описательного) плана, а ПД, ППЗ, ПО – модели процедурного типа (т.е. процедуры или алгоритмы).

Модель логики функционирования (МЛФ) СУД КА содержит информацию о режимах ее работы в НШС.

Модель контроля состояния (МКС) СУД КА отражает динамику изменения состояния ее подсистем.

Модель (алгоритм) идентификации (МИ) НШС представляет собой процедуру определения НШС по признакам.

Модель (алгоритм) парирования (МП) НШС реализует механизм синтеза сценария функционирования (последовательности операций), подлежащих исполнению СУД КА.

Уровень формализации, на котором реализуются алгоритмы, определяет характер процедур контроля текущего состояния СУД, диагностики и парирования НШС и отказов. На структурно-содержательном уровне моделирования им будут соответствовать механизмы волнового вывода на СС, на структурно-формальном – процессы изменения маркировки ВСП, а на формальном – алгоритмы логического вывода в продукционных системах.

Таблица 1

Система моделей и алгоритмов экспертной системы поддержки принятия решения по управлению движением КА в нештатных ситуациях

Уровень моделирования (формализации)	Вид модели (алгоритма) ЭСППР			
	Модель логики функционирования СУД КА (МЛФ)	Модель контроля состояния (МКС)	Модель (алгоритм) идентификации НШС (МИ)	Модель (алгоритм) парирования НШС (МП)
Содержательный	Содержательная модель логики функционирования СУД КА (СМФ)	Содержательная модель контроля состояния (СМКС)	Содержательная модель идентификации НШС (СМИ)	Содержательная модель парирования НШС (СМП)
Структурно-содержательный	Структурно-содержательная модель логики функционирования СУД КА (ССМФ)	Структурно-содержательная модель контроля состояния (ССМКС)	Структурно-содержательная модель идентификации НШС (ССМИ)	Структурно-содержа-тельная модель парирования НШС (ССМП)
Структурно-формальный	Структурно-формальная модель логики функционирования СУД КА (СФМФ)	Структурно-формальная модель контроля состояния (СФМКС)	Структурно-формальная модель идентификации НШС (СФМИ)	Структурно-формальная модель парирования НШС (СФМП)
Формальный	Формальная модель логики функционирования СУД КА (ФМФ)	Формальная модель контроля состояния (ФМКС)	Формальная модель идентификации НШС (ФМИ)	Формальная модель парирования НШС (ФМП)
Программный	Программная модель логики функционирования СУД КА (ПМФ)	Программная модель контроля состояния (ПМКС)	Программная модель идентификации НШС (ПМИ)	Программная модель парирования НШС (ПМП)

Технология представления знаний, используемая в рассматриваемой ЭСППР, получила название каузальной сети [7] – одной из разновидностей СС. Узлы сети (плоскости) представляют состояния СУД КА, а дуги каузальные отношения. СС имеет иерархическую структуру. Дуги сети, определяющие каузальные отношения, присоединяются к продукционным правилам с помощью указателей.

Плоскость парирования НШС является центральным звеном модели. Она представляет возможные действия по парированию НШС, которые можно объединить в четыре группы:

– переход в режим автономного управления;

– задействование циклограммы начального участка (переход в режим построения солнечной ориентации);

– переход в режим парирования НШС и задействование специальных режимов.

В плоскости диагностики НШС представлены результаты контроль состояния КА, который производится аппаратными (аппаратный контроль) или алгоритмическими (функциональный контроль) средствами.

При возникновении какой-либо из НШС, в телеметрической информации формируется признак, соответствующий происшедшему отказу (например, «Нарушение стабилизации на ДС», «Отказ СДП» и т.д.), или не формируется признак успеха проводимой операции (например, признак успешного проведения режима астрокоррекции).

В третьей плоскости – плоскости управления движением КА – представлены знания в виде способов и функциональных режимов управления движением КА.

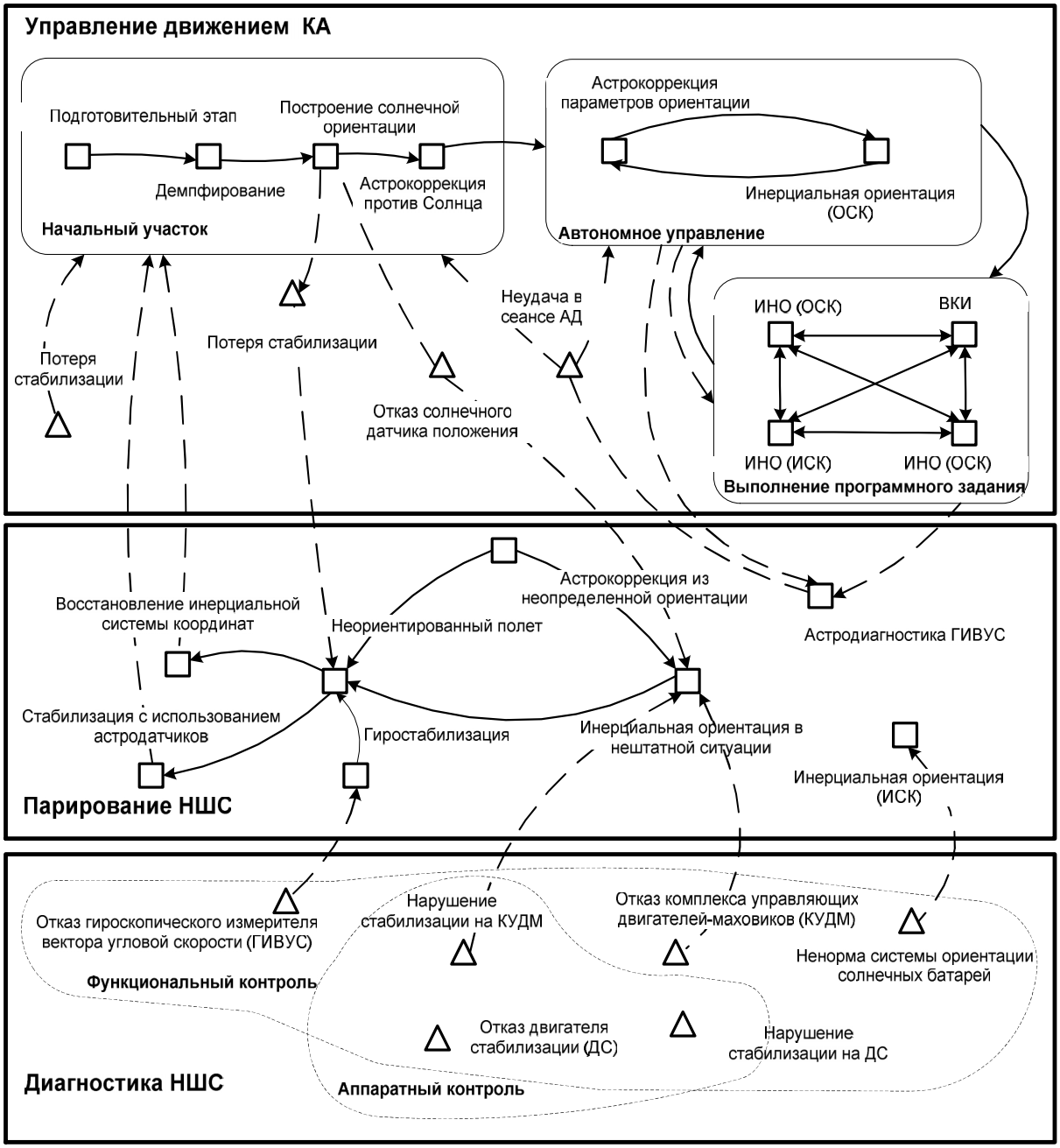


Рис. 3. Графическое представление каузальной сети

Вывод на семантической сети можно представить в продукционной системе, в которой каждая продукция имеет вид $Fr_1 \Rightarrow Fr_2$. Слева и справа в этих продукциях стоят фрагменты семантической сети. В зависимости от типа продукции она может описывать изменения в базе знаний или результаты промежуточных шагов в решателе. В базах знаний реализация продукции $Fr_1 \Rightarrow Fr_2$ называется процедурой поиска по образцу. В качестве образца при поиске выступает фрагмент Fr_1 . При этом в некоторых образцах для поиска могут присутствовать условия применимости продукций вида «Если имеет место Fr_1, то Fr_2, иначе Fr_3» [8].

Положительная сторона представления знаний семантическими сетями заключается в том, что это весьма простой и понятный способ описания на основании отношений между элементами (узлами и дугами). Однако с увеличением размеров сети существенно увеличивается время поиска по сравнению со способами, не имеющими стратегии. Кроме того, им присуща проблема гарантии пригодности результатов вывода, включающая также проблему наследования

свойств. И рассмотренное выше разделение сети является одним из подходов для ее решения. Требуется и такая реальная система как редактирование присоединенной процедуры в подсетях, используемое для проверки отсутствия противоречий с базой данных семантической сети. Альтернативой этому является способ описания семантической сети на метауровне.

Перспективные исследования в области представления знаний идут по пути вложения в семантические представления некоторых фрагментов процедурных и декларативных представлений с целью объединения их преимуществ комбинированном представлении.

Заключение

Современный уровень информационных технологий обеспечивает возможность высококачественной реализации всех компонентов ИСППР по управлению КА в условиях возникновения НШС. Построение комплекса моделей и алгоритмов ЭСППР является наиболее наукоемким и трудозатратным этапом создания данной системы, определяющим, в итоге, эффективность оперативного управления КА и качество его функционирования.

Применение системного подхода к созданию ЭСППР по управлению СУД КА позволило обосновать возможность и целесообразность их создания на основе современных информационных технологий.

Разработаны функциональные принципы построения и сформирован облик ЭСППР. Сформированы предложения по построению ЭСППР, а также выделены ее основные функциональные элементы.

Предложена технология формирования центральной системы ЭСППР – БЗ о процессе (логике) функционирования СУД КА, а также методика построения основных подсистем ЭСППР. На основе указанной методики сформирован состав системы моделей (алгоритмов) ЭСППР.

Предложено применение аппарата семантических сетей, продукционных систем для представления моделей (алгоритмов) ЭСППР на соответствующем уровне моделирования, позволяющие реализовать различные процедуры вывода в описанных моделях.

ЛИТЕРАТУРА

1. Геловани В.А. Интеллектуальные системы поддержки принятия решений в нештатных ситуациях с использованием информации о состоянии природной среды / В.А. Геловани, А.А. Башлыков, В.Б. Бритков, Е.Д. Вязилов. М.: Едиториал УРСС, 2015. 304 с.
2. Гудвин Г.К. Проектирование систем управления / Г.К. Гудвин, С.Ф. Гребе, М.Э. Сальгадо М.: Бином. Лаборатория знаний, 2016. 912 с.
3. Попов Э.В. Экспертные системы: Решение неформализованных задач в диалоге с ЭВМ. М.: Наука, 1987. 288 с.
4. Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. СПб.: Питер, 2000. 384 с.
5. Петрошенко А.В. Методы и модели инженерии знаний. Сетевые модели и методы обработки знаний. Вып. 1. СПб.: ВКА имени А.Ф. Можайского, 2004. 48 с.
6. Тейз А. Логический подход к искусственному интеллекту. От классической логики к логическому программированию / А.Тейз [и др.]. М.: Мир, 1990. 432 с.
7. Уэн Х. Представление и использование знаний / Х.Уэно, М ИСИдзука. – М.: Мир, 1989. 220 с.
8. Поспелов Д.А. Моделирование рассуждений. Опыт анализа мыслительных актов. М.: Радио и связь, 1989. 184 с.

E. N. Alyeshin, V. A. Tereshchuk, S. A. Tereshchuk (Military Space Academy name of A.F. Mozhaisky, Saint-Petersburg)

Model of knowledge representation of the expert decision support system for controlling the movement of the spacecraft in the event of abnormal situation

Of particular importance at the present time of space vehicles are the issues of ensuring the strengthening of maintenance of the selection processes of management decisions based on integrated automation, as well as increasing the efficiency of the functioning of the spacecraft in the event of abnormal situations. The urgency of solving these problems for space vehicles is caused, on the one hand, by the features of their functioning, and the other hand, by the lack of the possibility of timely control in the event of abnormal situations.

The use of traditional mathematical models and methods for describing the process of operation and control of spacecraft in abnormal situations is very difficult or impossible. In this connection, the category of knowledge of management specialists, whose intuition and experience are sometimes the only source of success of their actions, is brought to the forefront.

This article discusses the development and application of an expert system to support the decision-making on the movement control of a spacecraft in the event of abnormal situations. The place of intellectual decision support system in the automated control system for space vehicles, the composition of the expert decision support system for controlling the spacecraft motion control system and the designation of each of its elements, as well as the technology for forming a knowledge base containing information on the process (logic) of the functioning of the system motion control of the spacecraft.

Based on the principles of the system approach and modern information technologies, the composition of the models and algorithms for controlling the motion of the spacecraft in abnormal situations is described to build the knowledge base of the expert decision support system.

To represent the knowledge of expert at various levels of formalization (modeling) as models and algorithms that allow to implement different output procedures, a semantic network and a production system for the logic of the functioning of the spacecraft motion control system in abnormal situations are used.

АНАЛИЗ ЭВОЛЮЦИОНИРУЮЩИХ СПЕКТРОВ НЕСТАЦИОНАРНЫХ ПРОЦЕССОВ

А. В. Макшанов[1], **А. А. Мусаев**[2]
(*СПИИРАН, С.-Петербург, 199178, 14 линия ВО, 39, spiiran@iias.spb.ru*)

Аннотация

Предложена процедура анализа эволюционирующих спектров, допускающая относительно простую интерпретацию в привычных частотных терминах и позволяющая получать подробную диагностическую информацию с локализацией области возможных дефектов.

Ключевые слова: нестационарные процессы, эволюционирующие спектры, вобуляция, диагностика, авторегрессия.

Введение

В настоящее время имеется целый ряд подходов к проблеме формального обобщения процедур спектрального анализа на случай нестационарных процессов [2,6,7,810,14,17,19].

В представленном докладе рассмотрена специализированная процедура спектрально-временного анализа, используемая в задачах, связанных с построением образов (портретов) объектов по результатам измерений параметров их физических полей**.**

Задача изучена на примере анализа виброакустического сигнала, используемого для превентивной диагностики зарождающихся дефектов. При этом используются не только амплитуды основных частот возбуждения, но и их окраска, т.е. характер фазовой и амплитудной модуляции, соотношения между амплитудами гармоник, вид случайных выбросов в резонансных зонах и т.д.

Анализ характеристик нестационарных процессов в задачах диагностики технических систем

Как уже отмечалось, существует несколько подходов к проблеме формального обобщения процедур спектрального анализа на случай нестационарных процессов [2,6,7,810,14,17,19]. В частности, фирма Bruel & Kjaer три базовых алгоритма нестационарного спектрального анализа виброакустических сигналов на основе:

- преобразования Фурье на скользящем окне [15]:
- волнового преобразования (Wavelet Transform, [18]);
- распределения Вигнера-Вилля [2,7,19].

При критическом анализе таких подходов следует иметь в виду, что само утверждение о нестационарности несет лишь негативную нагрузку. Суть дела можно прояснить с помощью следующей простой статистической аналогии. Случайный вектор X с ковариационной матрицей S может быть записан в виде

$$X = CY \qquad (1)$$

где C - ортогональная матрица, а случайный вектор Y имеет диагональную ковариационную матрицу D, при этом

$$S = CDC'. \qquad (2)$$

Компоненты Y называют главными компонентами вектора X. Для стохастического процесса с нулевым средним аналогом (1) является представление $X(t) = \int f(t,w)dZ(w)$, а аналогом (2) - выражение автокорреляционной функции в виде

$$R(t,s) = \int f(t,w)\, f(s,w).dm(w),$$

где m(w) - интегрированная спектральная функция процесса X(t), при этом интерпретация аргумента w зависит от конкретного вида ядра f(t,w).

Пусть с другим случайным вектором X*, некоррелированным с X, связаны аналогичные представления X* = C*Y*, S* = C*D*C*'. Тогда если C = C*, то есть собственные векторы мат-

[1] Доктор технических наук, ведущий научный сотрудник.
[2] Доктор технических наук, главный научный сотрудник.

риц S и S* совпадают, то для произвольной линейной комбинации (a.X + a*.X*) имеют место представления

$$(a.X + a^*X^*) = C.(a.Y + a^*Y^*),$$
$$a^2S + a^{*2}S^* = C(a^2.D + a^{*2}D^*)C'.$$

Если C≠ C*, то ($a^2D + a^{*2}D^*$) уже не является ковариационной матрицей главных компонент вектора (aX + a*X*). Аналогично, если некоррелированный с X(t) процесс X*(t) имеет представление

$$X^*(t) = \int f(t,w)dZ^*(w)$$

с тем же самым ядром f(t,w), то произвольная линейная комбинация (aX+ a*X*) имеет такое же интегральное представление относительно процесса с независимыми приращениями (aZ(w)+ a*Z*(w)), а ее автокорреляционная функция имеет представление (2) со спектральной функцией вида $a^2m(w) + a^{*2}m^*(w)$. Известно, что стационарные в широком смысле случайные процессы порождаются общим ядром f(t,w) = exp(itw), что и обеспечивает наличие простых соотношений между спектральными плотностями их линейных комбинаций и, соответственно, между спектральными плотностями входных и выходных сигналов в линейных системах. В то же время для процессов с различными ядрами, как и для случайных векторов с различными матрицами C, подобные соотношения места не имеют. Тем не менее, имеется ряд конструкций ядер, которые позволяют получать эти соотношения для достаточно широких классов процессов в приближенном виде.

При построении обобщений спектрального анализа для нестационарных процессов естественно попытаться сохранить следующие свойства спектральной плотности:

- неотрицательность и аддитивность;
- однозначную связь с корреляционной функцией;
- связь с исходным рядом на основе преобразования Фурье;
- возможность выявления скачков;
- возможность определения передаточной функции для линейного преобразования;
- возможность оценивания по одной реализации;
- сводимость к обычной оценке спектральной плотности в случае процесса, стационарного в широком смысле.

В работе [17] формально доказано, что эта система требований является ***противоречивой.*** Различные варианты компромисса порождают целое семейство разнообразных подходов к определению спектральной плотности для различных классов нестационарных процессов. Эти подходы можно условно разделить на *три группы*.

Подходы первой группы исходят из определения локальной корреляционной функции и основаны, в конечном счете, на локальной аппроксимации корреляционной функции R(s,t) нестационарного процесса функцией одной переменной: R(s,t)≅r(s-t). Эти конструкции подробно изложены в [2]; в [8] они обсуждаются в связи с общим понятием нестационарной функции неопределенности. Их частными случаями являются упомянутые выше подходы на основе преобразования Фурье на скользящем окне и на основе аппроксимации Вигнера-Вилля.

Вторая группа представлена подходом Л.Б. Пристли [19] и относится к тем процессам, корреляционная функция которых может быть представлена в виде

$$R(s,t) = \int A(s,w)\overline{A(t,w)}\exp[iw(s-t)]dm(w),$$

где

$A(0,w)=1$, $A(t,w) = \int \exp(iq)dH(q,w)$

и |dH(q,w)| достигает максимума при q=0 для всех w. Для данного семейства функций F ={A(t,w)} определяются величины $B(w) = \int|q||dH(q,w)|$, $BF = [\sup BF(w)]^{-1}$ - *эффективная ширина данного семейства F.* Процесс x(t) называется *квазистационарным,* если для него BF отлично от нуля. Для такого процесса *эволюционирующий спектр* по отношению к семейству F определяется в виде $dF(t,w) = |A(t,w)|dm(w)$.

С этим определением связаны некоторые проблемы неоднозначности, но в целом такой подход представляется наиболее последовательным и свободным от внутренних противоречий. В качестве частных случаев он включает все три алгоритма, реализованных в [15], и дает при этом качественную характеризацию (в терминах B(w) и BF) классов процессов, для которых

каждый из этих алгоритмов является наиболее адекватным. Его подробное изложение вместе с соответствующими статистическими вычислительными процедурами приведено в [19, 20].

Третья группа основывается на представлении о том, что *общее утверждение относительно нестационарности некоего процесса не несет никакой позитивной нагрузки.* Их авторы подчеркивают прежде всего противоречия, связанные с попытками обобщить концепцию спектрального анализа на достаточно широкий класс нестационарных процессов и основываются на более детальных модельных предположениях, позволяющих описывать различные конкретные типы нестационарности. Принятие этой точки зрения ставит перед исследователем важную задачу: выделить тип нестационарности, наиболее характерный для выбранной сравнительно узкой предметной области, и использовать методы анализа, проблемно ориентированные на такой конкретный тип данных.

Подход Л.Б. Пристли является наиболее общим и позволяет рассмотреть с единой точки зрения все существующие конструкции. Его главный недостаток - отсутствие эффективной процедуры для корректного определения эффективной ширины данного локально-нестационарного процесса, которая, в конечном счете, определяется в том или ином варианте перебора. В этой связи в настоящее время наибольшее внимание привлекают процедуры *вейвелет-анализа* [18]**.** В этом варианте анализа рассматриваются различные функции (вейвелеты), зависящие от параметров сдвига и масштаба и образующие семейство с двойной ортогональностью - по сдвигу и по масштабу. Данному процессу ставится в соответствие матрица его ковариаций с каждой из функций данного семейства. Рассматриваются различные процедуры визуализации этой матрицы. Таким образом, вопрос о выборе эффективной ширины процесса отпадает, но получаемое представление обладает значительной избыточностью. Главный недостаток этого подхода - невозможность интерпретации результатов в привычных терминах спектрального анализа. Другим недостатком *вейвелет-анализа* является слишком большой объем необходимых вычислений.

В настоящее время основной акцент при исследованиях виброакустических процессов делается на анализе по огибающей. При этом возникает специальный тип нестационарности, теория которого детально исследована в [16]. Соответствующий анализ основан, как правило, на использовании преобразования Гильберта [2,8], однако, в особенности при анализе узкополосных сигналов, его можно успешно заменять вычислением скользящей оценки дисперсии с подходящим выбором весовой функции. Основная сложность здесь состоит в том, что при скользящем сглаживании коррелированных данных обычно обнаруживаются артефакты - ложные периодические составляющие с небольшими амплитудами (эффект Слуцкого-Юла), которые приходится специально отсеивать (например, усреднением результатов сглаживания с различными окнами). Аналогичная картина имеет место и для более сложных типов нестационарности.

В настоящее время в задачах цифрового спектрального анализа большое внимание уделяется получению спектральных оценок в параметрической форме [2, 3, 5, 6, 10, 12, 13, 19]. Они обладают целым рядом преимуществ перед непараметрическими оценками на основе БПФ, однако при работе в широких диапазонах частот обычно возникают модели слишком высоких порядков. В то же время такие оценки оказываются чрезвычайно эффективными при исследовании процессов вобуляции выделенных характерных частот, поскольку в достаточно узких спектральных диапазонах можно ограничиться моделями невысокого (обычно - до 4-го) порядка, которые можно идентифицировать на скользящем окне шириной 30-40 тактов измерения.

Предлагаемая схема анализа сводится к специально организованной процедуре линейного сглаживания на скользящем окне, принципы выбора ширины и формы которого подчиняются тем же закономерностям, что и для других схем скользящего сглаживания [4]. При этом возможно применение различных алгоритмов параметрического оценивания, систематически описанных в [2, 12]. При микропроцессорной реализации наиболее выгодно использовать оценки по МНК в рекуррентной форме на скользящем окне шириной l со сдвигом d [9]. Варьируя ширину окна l, можно выбрать ее оптимальное значение стандартными статистическими методами по критерию минимума дисперсии порождающего шума. Результаты анализа интерпретируются в терминах спектрально-временного описания Пристли [19]. Удается доказать, что такая интерпретация действительно правомочна для тех участков, на которых все корни характеристического полинома соответствующего разностного оператора оказываются мнимыми и удовлетворяют традиционным условиям устойчивости и обратимости. Разумеется, эта

схема адекватна только для специального частного типа нестационарности, однако имеющийся опыт обработки результатов виброакустических измерений позволяет утверждать, что именно такой тип нестационарности характерен для процессов развития неисправностей механического оборудования типа. Их традиционно фиксируемые проявления в спектре огибающей оказываются при этом лишь одним из сопутствующих эффектов. В рамках этой же техники возможно изучение и более сложных типов нестационарности, например, изменения во времени порядка модели и интенсивности порождающего шума [6].

Рассматриваются следующие варианты обработки и интерпретации результатов оценивания.

1. В спектре изучаемого сигнала выделяется достаточно узкий диапазон, содержащий отдельный пик и его боковые составляющие. Выраженные боковые составляющие обычно представляют собой артефакты, связанные с наличием на основной частоте модуляции по амплитуде. Оцениваются на скользящем окне параметры модели авторегрессии 2-го порядка. При нерекуррентном оценивании целесообразно использовать одно из стандартных сглаживающих окон (при экспериментах использовалось окно Бартлетта). Выводятся процессы вобуляции коэффициентов модели, оценки их спектральная плотность мощности (СПМ) и полученные по ним [2,3,5,12] процессы вобуляции СКО и центральной частоты. Основные периодические составляющие этих процессов удобно представлять в виде фигур Лиссажу на фазовой плоскости коэффициентов модели. В настоящее время накапливается опыт интерпретации полученных результатов в задачах технического диагностирования.

2. В спектре изучаемого процесса выделяется относительно широкий диапазон и на скользящем окне оцениваются параметры модели ARMA(p,q). Ширина окна должна быть достаточно большой, чтобы обеспечить разумную точность оценивания. По получившимся оценкам процессов изменения коэффициентов строятся оценки локальных СПМ. Из них собирается оценка эволюционирующей СПМ и ее линии уровня. На получающихся представлениях хорошо видны любые ударные воздействия и резонансы, которые они вызывают.

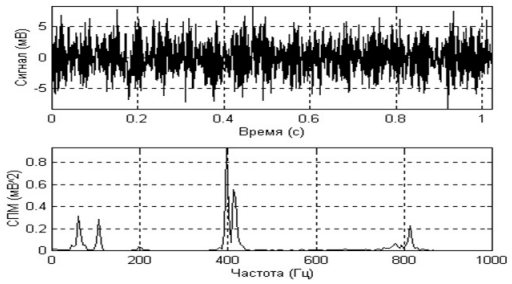


Рис. 1. Временной ряд наблюдений и его ПСМ

3. В обоих случаях возможно определение оптимальной ширины окна и оптимального порядка аппроксимирующей модели по критерию минимума дисперсии порождающего шума (критерий ФОП – финальной ошибки прогнозирования [2, 3, 5, 12]).

В качестве примера на рис. 1 приведен временной ряд показаний виброакселерометра, установленного на крышке кормового подшипника корабельного маршевого электродвигателя и оценка его СПМ, на рис.2-3 – представления процессов вобуляции в различных частотных диапазонах.

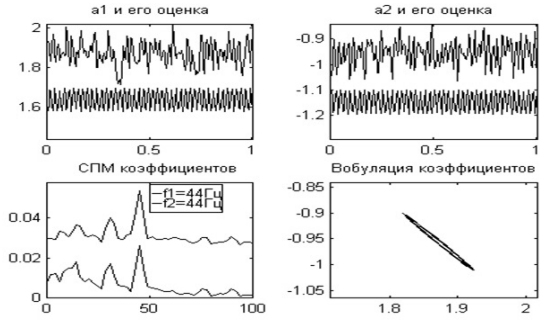


Рис.2. Вобуляция центральной частоты и амплитуды в диапазоне 200-400 Гц

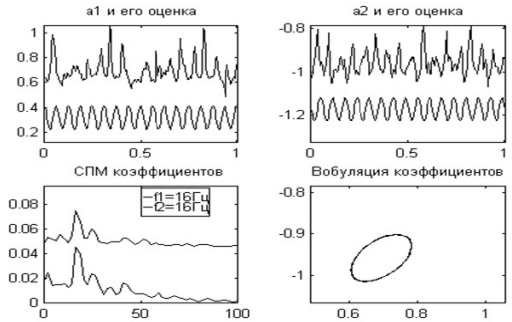


Рис.3. Представление вобуляций в диапазоне 200-400 Гц на плоскости коэффициентов модели AR(2)

На рис. 4-5 приведены оценки эволюционирующей СПМ в диапазоне 0-800 Гц и ее линии уровня в диапазоне 0-200 Гц. Использовалась аппроксимация на скользящем окне с параметрами l=100, d=30 и сглаживание окном Бартлетта.

Можно сделать вывод о том, что части двигателя, для которых характерно акустическое излучение со средней частотой около 70 Гц, испытывают апериодические ударные воздействия с резонансами на низких частотах. Варьируя изучаемый диапазон, можно локализовать неисправность, используя имеющуюся карту резонансных частот механизма.

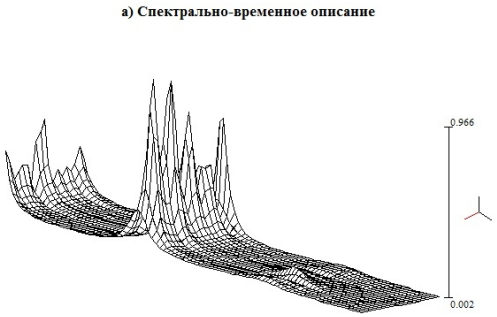


Рис.4. Представление вобуляций СПМ на плоскости “время-частота”: эволюция локальной СПМ в диапазоне 0-800 Гц, модель ARMA(10,0)

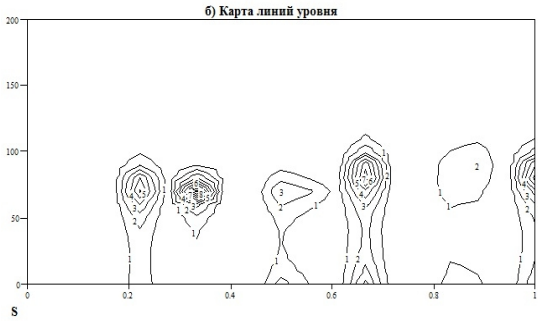


Рис.5. Представление вобуляций СПМ на плоскости “время-частота”: линии уровня локальной СПМ в диапазоне 0-200 Гц, модель ARMA(6,0).

Заключение

Предложенная организация процедуры анализа эволюционирующих спектров, в отличие от большинства существующих, допускает относительно простую интерпретацию в привычных терминах и позволяет получать подробную диагностическую информацию, локализованную по конкретным узлам изучаемого механизма.

Процедура позволяет строить достаточно детальные образы (портреты) объектов различной природы по результатам измерений параметров их физических полей и допускает естественные обобщения на случай многоканальных измерений, в том числе – различной физической природы.

Разработанный алгоритм имеет характер нелинейного сглаживания данных, результаты которой интерпретируются в терминах эволюционирующих спектров. По этой схеме удается построить аналогичные процедуры с другими вариантами интерпретации, проблемно-ориентированные на решение различных специальных классов задач.

Работа проводилась при поддержке гранта РФФИ 12–07–00302-а .

ЛИТЕРАТУРА

1. Батраков А.Ю., Макшанов А.В., Смирнов А.В. Адаптивный сглаживающий фильтр на скользящем окне для коррелированных шумов измерений // Вопросы радиоэлектроники. Сер. ОВР. 1988. №5. С.75-84.
2. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. — М.: Мир, 1989. 540с.
3. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Вып.1. — М.: Мир, 1974. 406с.
4. Гусак П.Л., Цыпкин Я. З. Теория сглаживания и ее применение // Измерение, контроль, автоматизация, 1988, 3. С.47-69.
5. Дженкинс Г., Ваттс Д. Спектральный анализ и его приложения. Т.1. М.: Мир, 1972. 318с.
6. Капустинскас С.Ю., Немура А.А. Идентификация линейных случайных процессов. Вильнюс: Мокслас, 1983. 160с.
7. Коэн Л. Время-частотные распределения. Обзор. // ТИИЭР. 1989. Т.77. № 10. С.72-120.
8. Макс Ж. Методы и техника обработки сигналов при физических измерениях. М.: Мир, 1983. Т.1. 312с.
9. Макшанов А. В.,Смирнов А. В., Шашкин А. К. Робастные методы обработки сигналов в радиотехнических системах синхронизации. СПб: Изд-во СПбГУ , 1991. 173с.
10. Макшанов А. В. Параметрический подход к задачам спектрально-временного описания нестационарных измерений навигационных параметров // Проблемы космической локации и навигации. СПб.: МО РФ, 1994. С. 94-100.
11. Макшанов А.В. Методы статистического анализа систем, представляемых временными рядами измерений, на основе рекуррентных алгоритмов с конечной памятью // Системный анализ при создании кораблей. 1995. Вып.4. С. 48-54.

12. Марпл С. Л.. Цифровой спектральный анализ и его приложения. М.: Мир, 1990. 265с.
13. Отнес Р., Эноксон Л. Прикладной анализ временных рядов. М.: Мир, 1982. 428с.
14. Цветков Э. И. Нестационарные случайные процессы и их анализ. М.: Энергия,1973. 128с.
15. Bruel & Kjaer. Non stationary signal analysis software type WT 9362. — System development, BU 0152, 1993.
16. Herbst L.J. Spectral analysis in the presense of variance fluctuations // J. of the Roy. Statist. Soc. Ser.B. 1964. V.21. №2. P. 354- 360.
17. Loynes R.M. On the concept of the spectrum for non-stationary processes // J. of the Roy. Statist. Soc. Ser.B. V.30. №1. P.204-237.
18. Meier Y. Wavelets: algorithms and applications. SIAM Rev. 1992. №36(3). P. 526-528.
19. Priestley M.B. Non-linear and non-stationary time series. London: Academic Press,1988. 227p.

A.V. Macshanov, A.A. Musaev (SPIIRAS, Saint-Petersburg)

Analysis of evolution spectra non-stationary processes

A procedure for analyzing evolving spectra is proposed. Advantages of the procedure are the possibility of interpreting the results in the usual frequency terms. The implementation of the procedure allows obtaining detailed diagnostic information with localization of the possible defects area.

ОПЫТ ФОРМИРОВАНИЯ ЭЛЕКТРОННОГО ОПРЕДЕЛЕНИЯ ИЗДЕЛИЯ НА ВСЕХ ЭТАПАХ ЕГО ЖИЗНЕННОГО ЦИКЛА НА ПРИМЕРЕ ДВУХСТЕПЕННОГО ПОПЛАВКОВОГО ГИРОСКОПА

А. Г. Баженов[1], **Б. Л. Шарыгин**[2]
(*АО "Концерн "ЦНИИ "Электроприбор",
С.-Петербург, ул. Малая Посадская, 30, office@eprib.ru*)

Аннотация

В докладе рассматривается опыт формирования информации об изделии на всех этапах его жизненного цикла на примере двухстепенного поплавкового гироскопа, способы организации работы с информацией, а также ее использования для повышения качества рассматриваемого изделия.

Ключевые слова: поплавковый гироскоп**,** жизненный цикл, качество.

Введение

Высокие показатели качества любого изделия достигаются путем применения различных подходов: использование современных средств проектирования и имитационного моделирования; применение передовых производственных технологий; обеспечение оптимальных методов сервисного обслуживания и т.д. Управление качеством – итерационный процесс, в основе которого содержится цикл Дёминга, и для его реализации необходимо соответствующее накопление информации, ее обработка и принятие корректирующих решений. Широкое внедрение компьютерных технологий на всех этапах жизненного цикла изделия позволяет организовать эффективное накопление и обработку собранной информации для более широкого круга изделий.

В докладе на примере двухстепенного поплавкового гироскопа (ДПГ) приведены новые результаты по формированию информации об изделии, полученной на этапах его проектирования, изготовления и эксплуатации. На различных объектах морского базирования сегодня функционирует более 200 систем инерциальной навигации и стабилизации (СИНС) типа «Ладога-М». Следовательно, эксплуатируются около тысячи ДПГ. Первые СИНС были переданы в эксплуатацию в 2003 году на экспортные фрегаты проекта 11356. В связи с расширением области применения СИНС постоянно возрастали требования к точностным и надежностным характеристикам. В результате ДПГ превратился в критичный элемент СИНС.

Жизненный цикл поплавкового гироскопа

Проектирование

Проектирование рассматриваемого в статье изделия было начато еще в 80-х годах XX-го века, и оно на протяжении всего времени постоянно изменялось с целью повышения эксплуатационных характеристик изделия, а также в связи с изменением номенклатуры используемых покупных изделий и материалов. В связи с бурным развитием компьютерных технологий в области систем автоматизированного проектирования (САПР), а также желанием применения средств компьютерного моделирования поведения элементов механических конструкций, были выполнены работы по перепроектированию изделия с использованием технологии 3D моделирования. В качестве САПР использовалась система Pro/ENGINEER фирмы PTC. Электроника разрабатывалась средствами Orcad фирмы Cadence. На рисунке 1представлена конструкторская модель изделия ДПГ , полученная с использованием САПР. Из рисунка видно, что модель выполнена с высокой степенью детализации, что диктуется необходимостью проведения точного компьютерного моделирования.

[1] Кандидат технических наук, начальник отдела интегрированных информационных систем.
[2] Кандидат технических наук, главный конструктор.

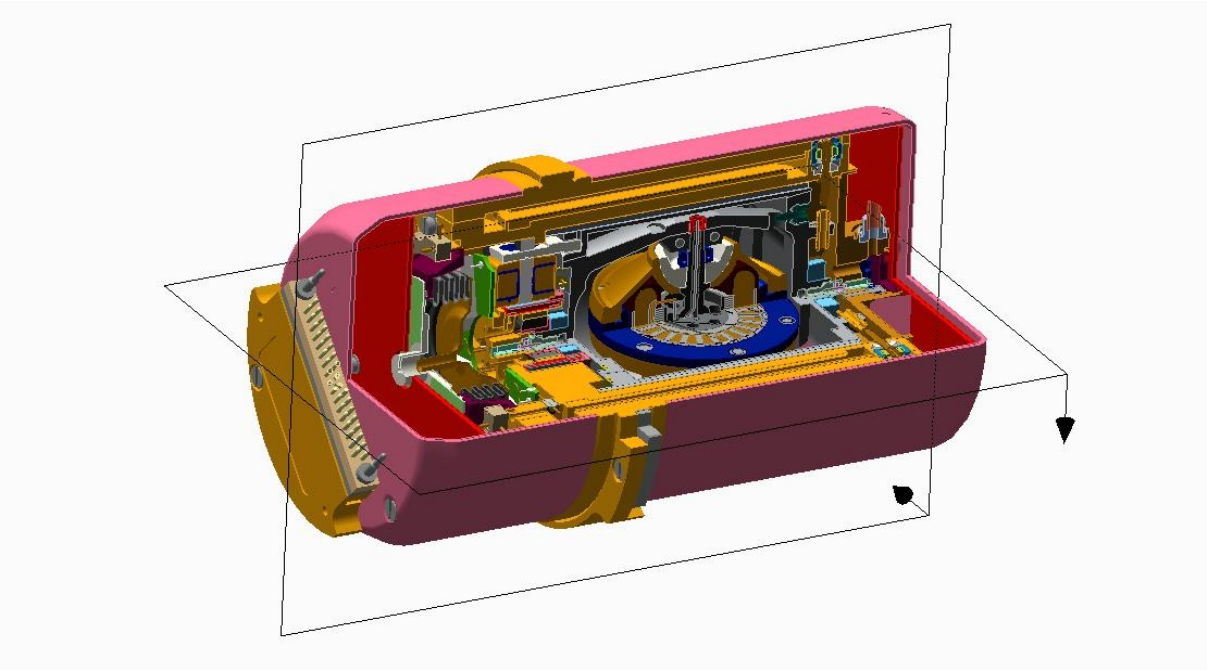


Рис. 1. Конструкторская трехмерная модель ДПГ

Инженерные расчеты

В 2011 году в стендовом зале института были зафиксированы несколько случаев перегрева ДПГ при регулировке СИНС типа «Ладога-М» [1].. Анализ данных таблиц, протоколов балансировки поплавковой камеры и контрольное изготовление позволили однозначно установить нарушения технологического процесса балансировки одним из исполнителей. Температура нулевой плавучести поплавковой камеры в гироскопе превышала номинальную на 2-3 градуса. Ужесточение контроля за балансировкой поплавковой камеры решили возникшую проблему и в течение последних семи лет обеспечивают температуру нулевой плавучести. Но ситуация с перегревами гироскопов спровоцировала тепловой расчет азимутального ДПГ [2]. В результате работы по расчету теплового режима азимутального ДПГ были проведены:

- расчет стационарного состояния прибора при температурах окружающей среды 45ºС и 48ºС;

- расчет нестационарного теплового режима прибора при его включении и прогреве при аналогичных температурах.

Результаты работы представлены в виде графической информации: распределений температур в сечениях, указанных в техническом задании, для стационарного режима; графиков изменения температуры во времени в определенных точках прибора. Приведены выводы по итогам расчетов и рекомендации по дальнейшим исследованиям.

Главная особенность прибора с точки зрения теплового режима состоит в том, что для поддержания требуемого уровня температур с точностью 0,1ºС используется собственная система термостабилизации, работу которой необходимо учитывать при проведении расчета. Из других особенностей стоит отметить нелинейный способ включения и прогрева прибора при изменяющейся температуре окружающей среды, использование в конструкции специфических материалов с достоверно неизвестными теплофизическими свойствами, микроскопические размеры некоторых элементов прибора. Для выполнения расчетов была использована геометрическая модель прибора ДПГ, полученная на этапе проектирования. Для решения задачи моделирования использовался метод конечных элементов и программные продукты на его основе: ANSYS, COMSOL и др. аналоги

Визуальное представление результатов расчета представлено на рисунке 2.

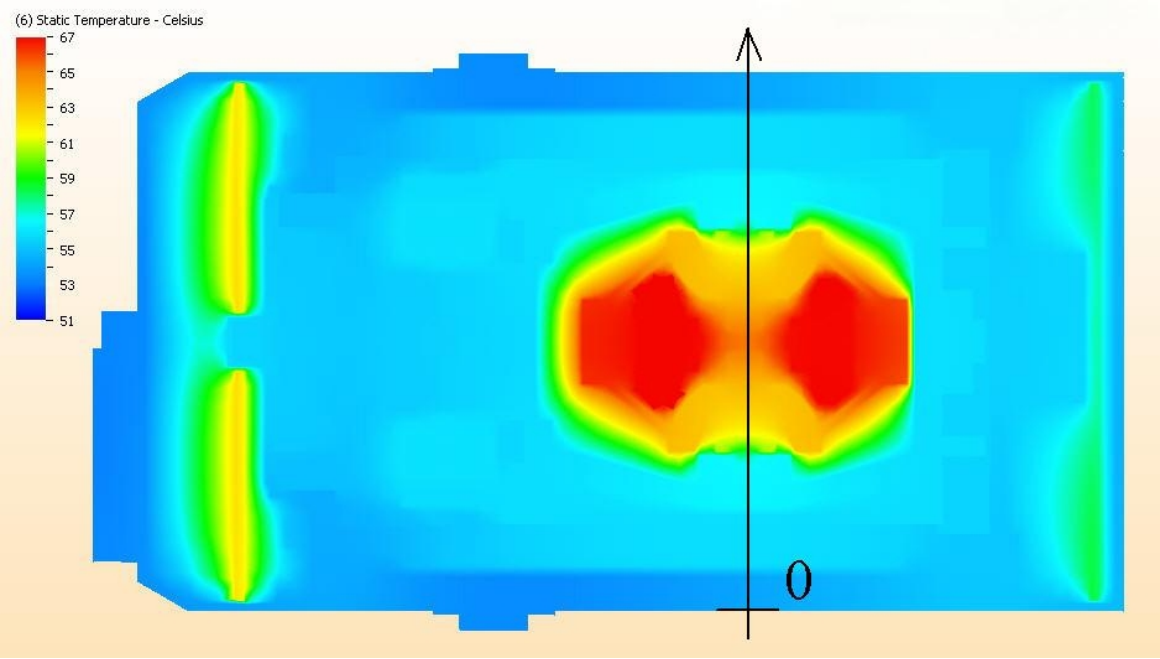


Рис.2 Визуальное представление результатов теплового расчета ДПГ

По результатам расчета сделаны следующие выводы:

- при включении гироскопа с температуры 20ºС прибор выходит на стационарный тепловой режим (изменение температуры в любой точке прибора не более 0,1ºС в течение 30 минут) примерно за 300 минут.
- разница между максимальной и минимальной температурами в приборе составляет приблизительно 16ºС;
- после включения гиромотора мощность, выделяемая им, компенсируется системой термостабилизации, но происходит перераспределение температур в поплавковой камере;
- отмечается значительный перепад температур между гиромотором и поверхностью поплавковой камеры (11ºС);
- использование материалов с высокой теплопроводностью при изготовлении элементов, расположенных в поплавковой камере, способствует выравниванию поля и снижению максимальных значений температур внутри неё.

Результаты теплового расчета были учтены при исследовании заклинивания гиромотора.

Кроме того был проведен анализ имеющийся информации о реализации осевых и радиальных зазоров в гироскопах выпуска 2009-2010 годов, когда заклинивания гиромотора практически не фиксировались. Сравнение статистик 2009-2010 годов и 2012-2014 годов выявило некоторые принципиальные изменения, хотя все находилось в установленных допусках. Пришлось вернуться к рассмотрению параметров газодинамической опоры и учесть тепловую деформацию ее элементов и раскрытие полусферических фланцев, обусловленное вращением роторной части гиромотора.

Была организована работа по определению температурной деформации в элементах газодинамической опоры[3]. При этом были решены следующие задачи:

- Разработка методики решения сопряженной задачи температурной деформации в элементах ГДО;
- Разработка расчетных моделей;
- Оценка влияния теплового поля на зазор между подвижным фланцем и полусферической опорой;
- Оценка влияния вращения ротора на зазор между подвижным фланцем и полусферической опорой;
- Анализ результатов.

В процессе этих работ [3], была создана и модель ротора с наложенной сеткой конечных элементов, приведенная на рисунке 3

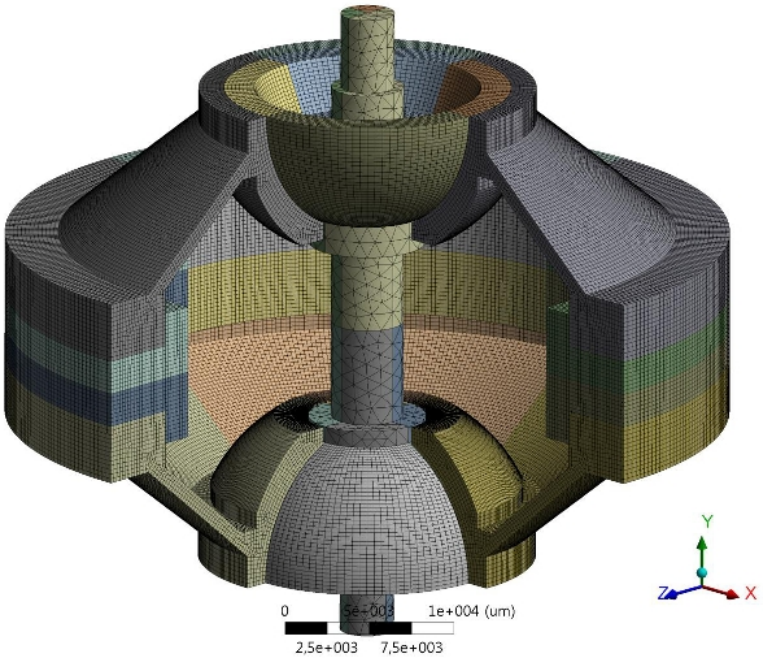


Рис. 3. Модель ротора с наложенной сеткой конечных элементов

Было промоделировано распределение температур в сечении поплавковой камеры (рис.4)

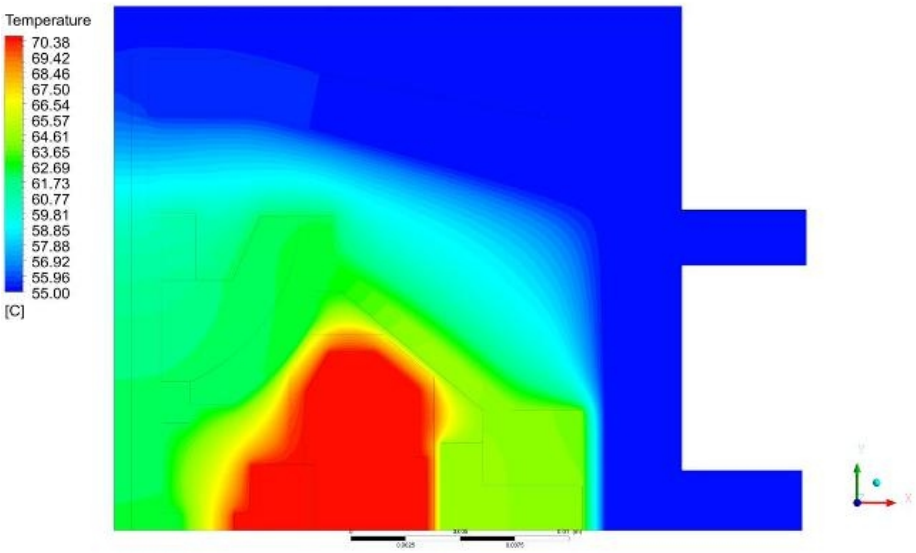


Рис. 4. Распределение температур в сечении поплавковой камеры

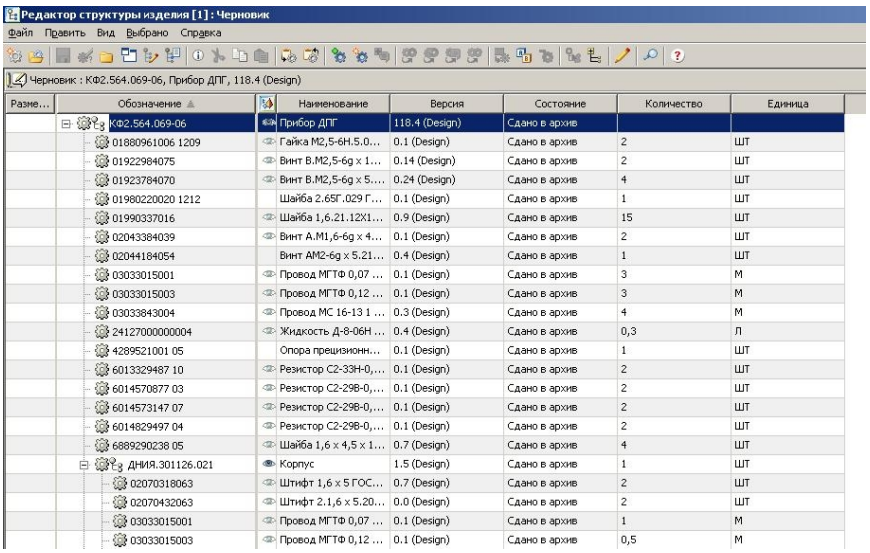
Получено уточнение максимальной температуры в гироскопе до 70 ºС, определено распределение уровня деформаций в элементах газодинамической опоры.

Оформление КД

Поскольку проектирование велось средствами САПР, то и результаты были получены в электронном виде. В концерне работа с электронной конструкторской документацией (КД) началась в 1998 году, когда были разработаны первые нормативные документы, описывающие указанный процесс. Первоначально для реализации электронного документооборота КД

использовался функционал операционной системы Windows NT 4.0, а также программное обеспечение собственной разработки, в том числе и для формирования спецификации в виде, позволяющем проводить ее машинную обработку [4]. Начиная с 2005 года в концерне внедряется система класса PDM, построенная на базе PLM Windchill фирмы PTC [5]. С публикацией государственных стандартов (ГОСТ 2.051, ГОСТ 2.053, ГОСТ 2.054) была проведена работа по гармонизации нормативных документов концерна с требованиями ГОСТ.

Действующая версия комплекта КД на ДПГ была оцифрована в 2007-2008 годах (в том числе разработаны трехмерные электронные модели). Указанный комплект размещен в PDM системе, где произведено его электронное согласование и сдача в архив организации. Также в PDM системе сформированы информационные объекты, описывающие конструктивную структуру изделия, что является краеугольным камнем при компьютерной обработке данных об изделии. На рисунке 5 показано визуальное представление объектов электронной структуры в PLM Windchill.



Разме…	Обозначение	Наименование	Версия	Состояние	Количество	Единица
	КФ2.564.069-06	Прибор ДПГ	118.4 (Design)	Сдано в архив		
	01880961006 1209	Гайка М2,5-6Н.5.0…	0.1 (Design)	Сдано в архив	2	ШТ
	01922984075	Винт В.М2,5-6g x 1…	0.14 (Design)	Сдано в архив	2	ШТ
	01923784070	Винт В.М2,5-6g x 5.…	0.24 (Design)	Сдано в архив	4	ШТ
	01980220020 1212	Шайба 2.65Г.029 Г…	0.1 (Design)	Сдано в архив	1	ШТ
	01990337016	Шайба 1,6.21.12Х1…	0.9 (Design)	Сдано в архив	15	ШТ
	02043384039	Винт А.М1,6-6g x 4…	0.1 (Design)	Сдано в архив	2	ШТ
	02044184054	Винт АМ2-6g x 5.21…	0.4 (Design)	Сдано в архив	1	ШТ
	03033015001	Провод МГТФ 0,07 …	0.1 (Design)	Сдано в архив	3	М
	03033015003	Провод МГТФ 0,12 …	0.1 (Design)	Сдано в архив	3	М
	03033843004	Провод МС 16-13 1 …	0.3 (Design)	Сдано в архив	4	М
	24127000000004	Жидкость Д-8-06Н …	0.4 (Design)	Сдано в архив	0,3	Л
	4289521001 05	Опора прецизионн…	0.1 (Design)	Сдано в архив	1	ШТ
	6013329487 10	Резистор С2-33Н-0,…	0.1 (Design)	Сдано в архив	2	ШТ
	6014570877 03	Резистор С2-29В-0,…	0.1 (Design)	Сдано в архив	2	ШТ
	6014573147 07	Резистор С2-29В-0,…	0.1 (Design)	Сдано в архив	2	ШТ
	6014829497 04	Резистор С2-29В-0,…	0.1 (Design)	Сдано в архив	2	ШТ
	6889290238 05	Шайба 1,6 x 4,5 x 1…	0.7 (Design)	Сдано в архив	4	ШТ
	ДНИЯ.301126.021	Корпус	1.5 (Design)	Сдано в архив	1	ШТ
	02070318063	Штифт 1,6 x 5 ГОС…	0.7 (Design)	Сдано в архив	2	ШТ
	02070432063	Штифт 2.1,6 x 5.20…	0.0 (Design)	Сдано в архив	2	ШТ
	03033015001	Провод МГТФ 0,07 …	0.1 (Design)	Сдано в архив	1	М
	03033015003	Провод МГТФ 0,12 …	0.1 (Design)	Сдано в архив	0,5	М

Рис. 5. Визуальное представление информационных объектов электронной структуры изделия в PLM Windchill

Изготовление

По влиянию на надежностные и точностные характеристики ДПГ на лидирующих позициях стоят:

- бериллиевые детали;
- поддерживающая жидкость типа Д8;
- заготовки для роторов из сплава Ю10Н15;
- прецизионные камневые опоры;
- золотая лента для токоподводов;
- магниты для датчика момента.

Названы всего 6 критичных позиций, а всего при изготовлении ДПГ используемая номенклатура элементов, материалов, сплавов, деталей составляет несколько сотен.

Требования к ним не только жесткие, но и всегда оригинальные. Объемы поставок незначительные, но они должны быть регулярными. Это не является особенностью ДПГ, это принципиальная особенность любого мелкосерийного гироскопического производства. Поэтому периодически и возникают проблемы в производстве с тем или иным комплектующим. В 2012 году выявилась негерметичность бериллиевых поплавковых камер, для которых бериллий поставлялся из Казахстана. Пришлось временно переориентироваться на американский бериллий.

В 2007 году прекратились поставки поддерживающей жидкости Д6 из-за прекращения выпуска по экологическим соображениям одного из компонентов этой жидкости. Пришлось переходить на жидкость Д8 с несколько иным удельным весом. Пришлось дорабатывать поплавковую камеру, чтобы уменьшить ее вес на 7 грамм.

Прецизионные камневые опоры для нас и большинства других изготовителей поплавковых гироскопов со времен Советского Союза поставлял Петергофский Часовой завод. Но и он не выдержал реалий 90-ых и нулевых годов и был вынужден отказать нам в поставках 5-6 лет назад. Пришлось переориентироваться на Арзамасский приборостроительный завод.

Сплав типа АЛНИ Ю10Н15 для роторов ДПГ был разработан в нашем институте и до 2004 года отливался в нашей организации. После закрытия нашего литейного производства заготовки поставлял завод Красный Октябрь. Конечно, были проблемы. Но с 2012 года перешли на поставки заготовок с Владимирского НПО «Магнетон». Проблемы тоже были и будут. Для исследования роторов из проблемных гироскопов привлекли растровый электронный микроскоп и рентгеноспектральный микроанализ [6]. Проведенные исследования химического состава магнитотвердого сплава на уровне заготовок и роторов позволили уточнить соответствие химического состава сплава техническим требованиям по содержанию основных и легирующих элементов Ni, Al, Cr и V. С помощью однофакторного дисперсионного анализа проведено исследование гомогенности химического состава роторов. Получены электронные изображения отдельных областей поверхностей роторов, которые позволяют оценить качество финишной обработки роторов при изготовлении. Проведен анализ обнаруженных включений, в том числе сульфидов и фторидов. Обязали поставщика заготовок роторов проводить анализ химического состава заготовок. Начали совместную работу по снижению серных включений.

Изделие ДПГ характеризуется сотнями параметров, которые влияют на конечное качества продукта. С целью оценки их влияния, был организован сбор их значений в процессе производства на следующих этапах:

- на определенных этапах технологического процесса во время изготовления и сборки, а также в процессе регулировки;
- в процессе регулировки в составе СИНС;
- в процессе регулировки в составе навигационного комплекса.

Зафиксированные параметры изготовления каждого экземпляра изделия хранятся в специализированной базе данных, которая фактически представляет собой технологический электронный паспорт изделия. Для реализации процессов формирования технологического паспорта изделия, а также анализа зафиксированных значений параметров, была использована информационная система класса MES – Wonderware, разработанная фирмой Schneider Electric [7].

Эксплуатация

Как известно ДПГ установлены в СИНС типа «Ладога-М», которых изготовлено и передано в эксплуатацию более 200 систем. Системы установлены на надводные корабли, подводные лодки и некоторые специальные суда. Кроме кораблей нашего ВМФ, СИНС эксплуатируются и в ВМС Китая, Индии, Вьетнама и т.д. Но для большинства ДПГ эксплуатация начинается в стендовом зале института и продолжается от нескольких сотен часов до 1.5-2 лет. Этот период эксплуатации наиболее ответственный, но и наблюдаемый. Он не только наблюдаемый, но и документируемый с помощью комплекта испытательной аппаратуры СИНС.

При оформлении отказа, при исследовании отказа ДПГ эта информация всегда используется и попадает в соответствующие итоговые документы.

Службы отдела технического контроля ведут накопление информации по отказам, рекламациям. Эта информация доступна для работников организации в виде электронной таблицы.

По основному периоду эксплуатации такой информации практически нет. Исключение составляют гироскопы, которые отказали на объектах. Вместе с отказавшим элементом поступает акт отказа или рекламация, которые несут много полезной информации об обстоятельствах отказа.

Анализ информации, получаемой в процессе жизненного цикла изделия

В процессе жизненного цикла гироскопа накапливается следующая информация:

- карты измерений деталей;
- протоколы отдельных технологических операций;
- таблицы параметров при изготовлении, регулировке и испытаниях узлов и гироскопа;

- пуски гироскопов в ОП-99;
- акты сходов;
- акты исследования сходов;
- программы и протоколы исследования сходов
- пуски изделий типа «Ладога-М»;
- акты отказов гироскопов;
- программы и протоколы исследования отказов;
- рекламации входящие;
- акты исследования отказов;
- акты исследования рекламаций;
- решения;
- карты разрешения;
- акты и протоколы периодических испытаний;
- акты и протоколы типовых испытаний.

Подавляющее большинство этих документов имеется в электронном виде и соответствующем формате. Но необходимо иметь эту информацию в электронном виде, каталогизировать ее с привязкой к заводскому номеру гироскопа и обеспечить открытый доступ к ней заинтересованных сотрудников.

За последние несколько лет при анализе информации гироскопов началось использование корреляционного и дисперсионного анализов [8]

Заключение

Организация сбора максимально полной информации о сложном изделии на всех этапах его жизненного цикла в виде, позволяющем провести ее анализ по различным критериям, позволяет быстро добиться требуемых показателей качества изделия. Компьютерные технологии на сегодняшний день позволяют формировать базы данных, содержащие информацию об изделии, а развитие технологий работы с «большими данными», организовать их обработку.

ЛИТЕРАТУРА

1. Пешехонов В.Г., Миронов Ю.В., Шарыгин Б.Л. Единая система инерции-альной навигации и стабилизации “Ладога-М” // Морская радиоэлектро-ника. – 2003 г - №1(4).
2. Громов Д.С. Безмен В.С., Евстифеев М.И. Отчет о выполнении расчета теплового режима двухстепенного поплавкового гироскопа Е32-148 КФ2.562.148-01 ЦНИИ Электроприбор 2013 г
3. Беляев Я.А. Бочаров А Отчет с результатами работ по теме «Решение сопряженной задачи температурной деформации в элементах ГДО ЦНИИ Электроприбор 2016 г
4. Баженов А.Г., Безмен В.С., Гутнер И.Е., Шарыгин Б.Л. Развитие САПР в ЦНИИ "Электроприбор" в российский период. В сборнике: Информационные технологии в управлении (ИТУ-2016) Материалы 9-й конференции по проблемам управления. Председатель президиума мультиконференции В. Г. Пешехонов. 2016. С. 507-514.
5. Баженов А.Г., Донецкая Ю.В Автоматизация процесса разработки и согласования конструкторской документации средствами системы Windchill. Гироскопия и навигация. 2007. № 2 (57). С. 107.
6. А. В. Старцева, Б. Л. Шарыгин, Е. Г. Литуненко Электронный микроскоп при исследовании проблем в мелкосерийном производстве ДПГ . Доклад на XX конференции молодых ученых. 2018 г
7. Буцык А.Я., Донецкая Ю.В., Шарыгин Б.Л. Электронный технологический паспорт. Известия вузов. Приборостроение, № 3(60) , 2017 г
8. Шарыгин Б.Л. Корреляционный анализ при исследовании проблем в мелкосерийном производстве двухстепенных поплавковых гироскопов . Известия вузов. Приборостроение На рассмотрении.

ФОРМАЛЬНАЯ МОДЕЛЬ И МЕТОДЫ ОПИСАНИЯ СИСТЕМ СО СЛОЖНЫМ ПОВЕДЕНИЕМ НА ПРИНЦИПАХ ОБОБЩЁННЫХ АВТОМАТОВ

И. В. Афанасьева[1]
(*САО РАН, 369167, Нижний Архыз, КЧР; Университет ИТМО, 197101, С.-Петербург, Кронверкский пр, 49, тел. +7 878 229-34-32, riv@sao.ru*)

Д. К. Левоневский[2]
(*Санкт-Петербургский институт информатики и автоматизации РАН, 199178, С.-Петербург, 14-я линия Васильевского острова, 39, +7 812 328-43-69, dl@iias.spb.su*)

Ф. А. Новиков[3]
(*Санкт-Петербургский политехнический университет Петра Великого, С.-Петербург, ул. Политехническая, 29, тел. +7 812 552-65-08, fedornovikov51@gmail.com*)

Л. Н. Федорченко[4]
(*Санкт-Петербургский институт информатики и автоматизации РАН, +7 812 328-19-19, lnf@iias.spb.su*)

Аннотация

Сложные событийно-управляемые программные системы постоянно должны реагировать на внешние и внутренние воздействия (стимулы). Ответная реакция таких систем зависит от их предыдущего поведения, которое обычно определяется с помощью автоматных моделей, основанных на графах переходов состояний. Для описания графов переходов состояний используются специальные языковые средства, как графические, так и текстовые. Автоматные модели наиболее удобны для проектирования систем со сложным поведением, но имеют ряд ограничений: отсутствие поддержки асинхронности и параллельности. Предлагаемый доклад посвящён описанию языка программирования высокого уровня CIAO, из программы которого генерируется модель системы автоматов, симулирующих функционирование исходной реагирующей системы.

Ключевые слова: асинхронные параллельные реагирующие системы, грамматика в регулярной форме, граф переходов состояний.

Введение

Большой класс технологических инструментов для производства программ обработки данных использует принцип синтаксического управления. В типичных приложениях, таких как трансляция языков программирования, этот принцип давно и успешно используется. Управление процессом трансляции определяется синтаксической структурой предложений входного языка.

Особенность таких программ состоит в том, что структура входных данных (входного языка) и система управления обработкой данных такой программы определяется сходным образом: одной и той же формальной грамматикой, или системой конечных автоматов, или же управляющей таблицей. Главная часть логики такой программы сосредоточена в схеме взаимодействующих автоматов, а сама программа обращается к ней на каждом шагу её работы. Управляющая таблица, в частности, определяет порядок вызова семантических процедур, осуществляющих обработку данных в процессе анализа входного текста.

Цель исследования: разработать методику автоматизированного формирования событийно-управляемой программной системы, с помощью языка автоматного программирования CIAO (Cooperative Interaction of Automata Objects) [1–3] для проектирования асинхронных параллельных реагирующих систем.

[1] Ведущий инженер, аспирант
[2] Научный сотрудник
[3] Доктор технических наук, профессор
[4] Старший научный сотрудник, доцент

Краткая историческая справка

Данная тема научных исследований была поставлена с целью уменьшить потери из-за сбоев и недостаточной производительности специального программного обеспечения сбора и обработки данных от астрономических фотоприёмных устройств (СПО ФПУ), а также необходимостью повышения продуктивности разработок и дальнейшего сопровождения программного обеспечения.

Методы разработки сложных программных систем определяются используемой моделью поведения. В известной классификации поведенческих моделей программ Д. Харела выделяются *реагирующие* и *трансформационные системы* [4]. Большинство систем управления техническими средствами являются реагирующими.

Применяемое СПО ФПУ взаимодействует с внешней средой посредством обмена сообщениями [2]. СПО ФПУ может выполнять одновременно несколько независимых операций, поэтому оно рассматривается как параллельное приложение. Эффективность достигается за счёт использования компьютера с несколькими или многоядерными процессорами, которые получили в настоящее время широкое распространение. Сложность использования многоядерных и многопроцессорных структур заключается в необходимости применения специальных средств программирования, так как первоначально программы не являются параллельным представлением алгоритмов решения задач.

С целью преодоления этих трудностей создан графический язык CIAO [3], который предназначен прежде всего для:

1) наглядного описания и визуализации систем со сложным поведением;
2) реализации взаимодействия параллельных программных объектов;
3) построения моделей систем со сложным поведением и доказательного исследования и верификации этих моделей;
4) быстрого создания прототипов образцов параллельных реагирующих систем.

Язык CIAO показал хорошие результаты применительно к СПО ФПУ, в связи с чем возникла идея расширить область применения этого языка. Далее будем рассматривать посылку и приём сигналов как абстрактные операции, которые выполняются системой взаимодействующих параллельных автоматов. В реализации возможны вариации в зависимости от области применения. Можно, например, считать, что сигналы передаются и обрабатываются мгновенно, а можно учитывать скорость передачи и время, необходимое для перехода автомата из одного устойчивого состояния в другое.

Отсюда и интерес к языку CIAO, на котором хотелось бы выразить программы для сети взаимодействующих автоматных объектов, но при этом может понадобиться определённое расширение языка за счёт добавления в него новых конструкций, специфических для конкретных предметных областей.

Применительно к СПО ФПУ язык используется преимущественно в графической форме, поскольку наглядность описания является доминирующим требованием. Графическая часть языка CIAO является совместной комбинацией диаграммы классов, диаграммы автомата и диаграммы компонентов унифицированного языка моделирования UML, расширенных понятиями требуемого и предоставляемого интерфейсов автомата.

Для применения в широком классе предметных областей с возможностью настройки языка на предметную область (кастомизация) предпочтительнее текстовая форма языка, поскольку она открывает возможность построения синтаксически управляемых программ обработки входных структурированных данных (процессоров), о которых шла речь выше. В качестве перспективной формальной базы для построения процессоров разработана методика, основанная на использовании грамматик с обобщёнными регулярными выражениями и атрибутами в виде специальных символов, семантик, для представления контекстных ограничений в синтаксических граф-схемах.

Формальными моделями методики построения языковых процессоров, моделирующих кооперативное взаимодействие автоматных объектов, являются регулярные выражения – как средство описания языка, конечный автомат – в качестве адекватного средства его обработки и его обобщения в виде диаграмм переходов состояний.

Теоретическим обоснованием этих моделей являются теорема С.Клини и её следствия о том, что класс регулярных множеств является минимальным классом, содержащим все конечные множества, замкнутым относительно операций объединения, конкатенации и замыкания [9].

Регулярные множества распознаются конечными автоматами и представляются регулярными выражениями.

Использование этой модели ограничено её применением лишь к регулярным языкам, поэтому модель обобщена до класса контекстно-свободных языков в регулярной форме (КСР-языков) с использованием специальных автоматов [5]. КСР-язык порождается КСР-грамматикой – обобщением контекстно-свободной грамматики (КС-грамматики).

Основные определения даны в работах [6–8]. Напомним их неформально.

Как известно, формальным языком является набор цепочек языковых токенов (лексем или терминалов) в соответствии с грамматикой языка [9]. Грамматика языка состоит из контекстно-свободной части в виде КС-правил и контекстно-зависимой части (ограничений). КС-правила представлены в форме с обобщёнными регулярными выражениями. Обобщение конечно-автоматной модели обработки языков сводится к следующему:

- вводится итерация, обозначаемая знаком «диез» (#) или обобщённая итерация, или итерация Г.С. Цейтина, которая не расширяет множество регулярных слов и может быть определена через традиционную (одноместную) операцию Клини (*) как (P#Q) = P, (Q, P)*. Она удобна при работе со стеком и частично решает задачу минимизации регулярного выражения по числу вхождений символов из объединённого алфавита всех символов грамматики;
- язык описывается с помощью КСР-грамматики – обобщения КС-грамматики. Класс языков не расширяет, но снимает лишнюю структурированность грамматического описания языка, в частности, простые конструкции, например, последовательности, списки с разделителями описываются без рекурсивных правил, только с применением итерации;
- строится синтаксическая граф-схема (СГС) – графический аналог КСР-грамматики, стартовый объект для синтеза распознавателя (анализатора) языка.

Грамматическое описание языка CIAO

Правила грамматики записываются следующим образом:

Нетерминал : Регулярное_выражение .

где Нетерминал – одно из вышеперечисленных обозначений для нетерминалов, а Регулярное_выражение задается следующим синтаксисом в формализме Наура-Бэкуса:

```
Регулярное_выражение ::= {
        Пусто | Лексема | Нетерминал | Семантика |          //Базовые элементы (1)
    Регулярное_выражение1   Регулярное_выражение2 |        //Конкатенация (2)
    Регулярное_выражение1 ';' Регулярное_выражение2 |      //Альтернативный выбор (3)
    Регулярное_выражение1 '#' Регулярное_выражение2 |      //Итерация (4)
                '[' Регулярное_выражение ']'               //Необязательный элемент (5)
                '(' Регулярное_выражение ')' }.            //Выражение в скобках (6)
```

В фигурных скобках через вертикальную черту перечислены альтернативы. В строке (1) перечислены базовые элементы, составляющие регулярное выражение: Пусто – пустое выражение (отсутствие чего-либо), Лексема, Нетерминал и Семантика. Строка (2) представляет операцию конкатенации, не имеющую специального знака для своего обозначения; строка (3) – это операция альтернативного выбора, знаком которой является точка с запятой; строка (4) задает операцию итерации, знаком которой является диез, строка (5) задаёт необязательную конструкцию, то есть [P] = (P ;), а строка (6) позволяет заключать регулярное выражение в круглые скобки, чтобы рассматривать его как один операнд в операциях конкатенации, альтернативного выбора и итерации. Для обозначения комментария используется комбинация символов //.

Ниже описан конкретный синтаксис языка CIAO с минимумом разделителей. Ради краткости считается, что тип идентификаторов определяется на лексическом уровне. Интенсивно используется итерация, рекурсивные правила в явно выписанной части грамматики не используются.

Нетерминалы

В текущей версии грамматики языка CIAO выделяются следующие 15 нетерминалов для обозначения отдельных языковых структур: P (описание одного автоматного объекта на языке CIAO), N (номер и имя автоматного объекта), V (раздел переменных), E (раздел событий – входных команд), A (раздел действий – выходных команд), R (раздел выходных запросов), Q (раздел входных запросов), U (раздел устойчивых состояний), W (раздел неустойчивых состояний), D (раздел состояний выбора), G (раздел групповых состояний), L (раздел внешних связей), B (булевское выражение), S (общий нетерминал для всех видов состояний), Action (список действий и/или выражений на переходе). Начальным нетерминалом, из которого порождается текст любой синтаксически правильной программы, является нетерминал P. Полная программа на CIAO является последовательностью разделов в фиксированном порядке.

Терминалы

Лексический анализатор преобразует цепочку символов из входного файла в последовательность лексем: несколько лексем общего вида – '<char>' (литера), '<number>' (целое), '<operation>' (знак операции), '<string>' (строка), '<tag>' (идентификатор) – и ряд "частных" лексем – ключевых слов, перечень которых определяется грамматикой входного языка CIAO. Лексемы выступают в роли терминалов и заключены в кавычки. В языке CIAO можно выделить следующие группы лексем:

1. Лексемы–названия разделов:
'VAR', 'EVENT', 'ACTION', 'REQUEST', 'QUERY', 'STATE', 'WAGGLY', 'DECISION', 'GROUP', 'LINK';

2. Лексемы общего вида (в их обозначениях присутствуют угловые скобки < и >):

- '<p_num>' – уникальный номер автоматного объекта в пространстве имён, записывается так: **#**целое;
- '<p_nm>' – имя автоматного объекта, один или несколько произвольных идентификаторов;
- '<v_nm>' – имя локальной переменной, произвольный идентификатор;
- '<type>' – идентификатор встроенного типа, то есть один из идентификаторов 'string', 'Real', 'Integer', 'Boolean', или идентификатор внешней структуры (класса);
- '<expr>' – логическое или арифметическое выражение;
- '<e_nm>' – имя события – входной команды (часть предоставляемого интерфейса), произвольный идентификатор;
- '<a_nm>' – имя действия – выходной команды (часть требуемого интерфейса), произвольный идентификатор;
- '<r_nm>' – имя выходного запроса (часть требуемого интерфейса), произвольный идентификатор;
- '<q_nm>' – имя входного запроса (часть предоставляемого интерфейса), произвольный идентификатор;
- '<u_nm>' – имя устойчивого состояния, произвольный идентификатор;
- '<w_nm>' – имя неустойчивого состояния, произвольный идентификатор;
- '<d_nm>' – имя состояния выбора, идентификатор, записывается так: **dcsn**целое. Под целым числом подразумевается номер, указанный в состоянии;
- '<g_nm>' – имя группового состояния, произвольный идентификатор, записывается так: **grp**целое. Под целым числом подразумевается номер, указанный в состоянии;
- '<f_num>' – номер вызывающего автоматного объекта, которому данный автомат предоставляет указанный интерфейс, записывается так: **#**целое;
- '<t_num>' – номер вызываемого автоматного объекта, чей указанный интерфейс требуется, записывается так: **#**целое;
- '<val>' – самоопределённое значение: либо “строка”, либо истинностное значение 'T', 'F', либо натуральное число, либо число с плавающей точкой;
- лексемы-метасимволы языка, изображаемые одним или двумя специальными символами:
- ':' – после двоеточия указывается тип определяемой переменной или возвращаемого значения, для входного запроса указывается имя локальной переменной;

- `','` – соединяет последовательно;
- `'|'` – разделяет альтернативы;
- `'[' ']'` – выделяет сторожевое условие на переходе;
- `'(' ')'` – группирует;
- `'='` – оператор присвоения значения;
- `'->'` – признак начала или окончания перехода;
- `'/'` – признак начала действий на переходе.

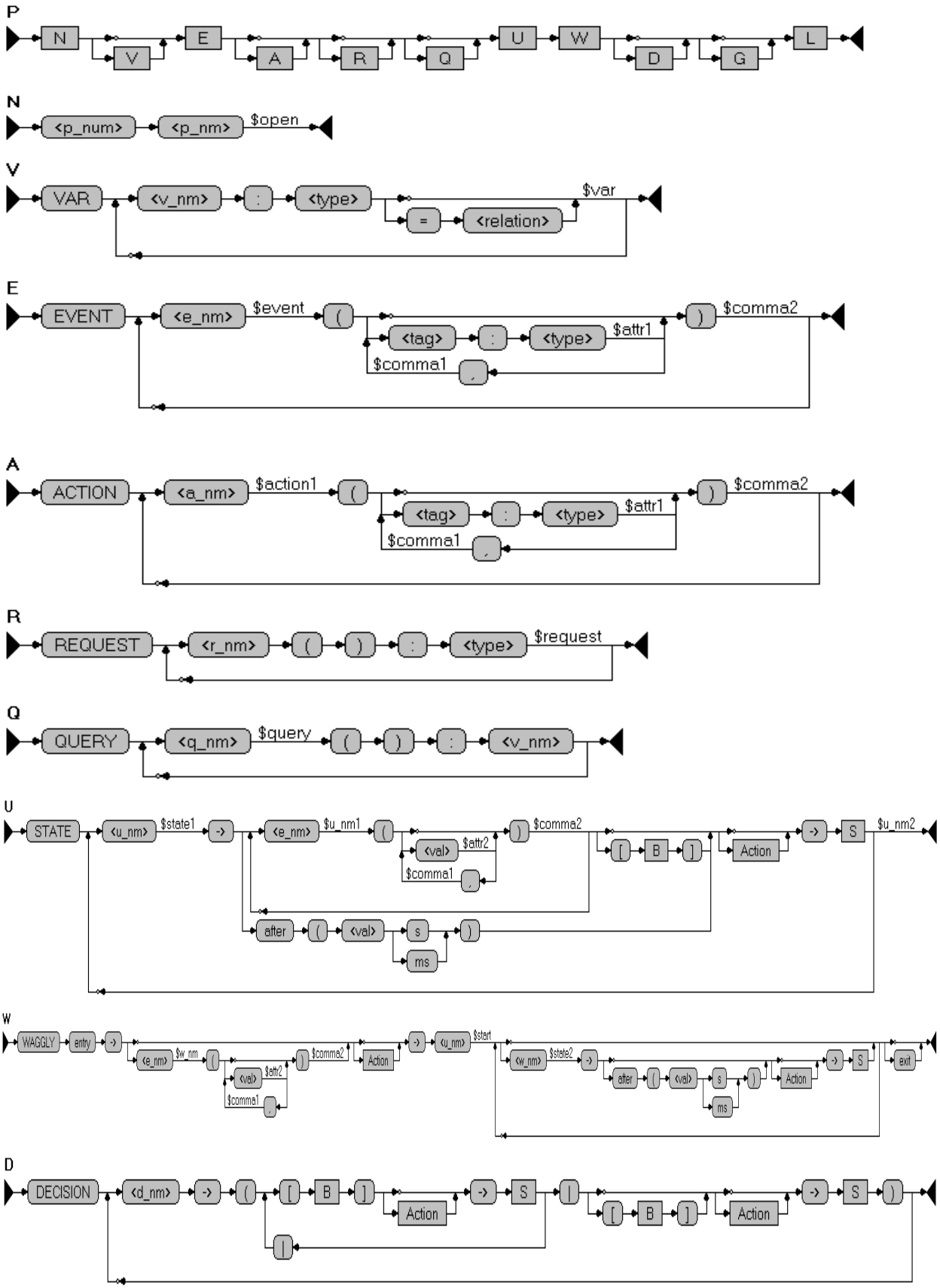
3. Лексемы – зарезервированные слова:
`'Real'` – тип число с плавающей точкой, `'Integer'` – целочисленный тип, `'Boolean'` – булевский тип, `'after'` – выход из состояния по прошествии указанного интервала времени, `'s'` – размерность в секундах, `'ms'` – размерность в миллисекундах, `'entry'` – начальное состояние, `'exit'` – заключительное состояние, `'T'` – значение “истина” для булевского типа, `'F'`– значение “ложь” для булевского типа.

Для учета неформализованных контекстных зависимостей в правила грамматики в виде регулярных выражений введены следующие **39** семантик – процедур, которые исполняются, если в процессе распознавания входного текста очередная распознанная лексема является той, которая в грамматическом правиле следует за данной семантикой; причем эта семантика имеет доступ ко всем параметрам данной лексемы; название каждой семантики начинается со знака "`$`": `$after`, `$action1`, `$action2`, `$action3` `$action4`, `$attr1`, `$attr2`, `$attr3`, `$close`, `$comma1`, `$comma2`, `$cond1`, `$cond2`, `$d_nm1`, `$d_nm2`, `$else`, `$event`, `$g_nm`, `$link1`, `$link2`, `$open`, `$query`, `$r_nm`, `$request`, `$sign1`, `$sign2`, `$sign3`, `$sign4`, `$start`, `$state1`, `$state2`, `$state3`, `$state4`, `$state5`, `$state6`, `$u_nm1`, `$u_nm2`, `$var`, `$w_nm`.

Правила грамматики языка CIAO:

```
P : N [V] E [A] [R] [Q] U W [D] [G] L .
N : '<p_num>' '<p_nm>' $open .
V : 'VAR' ( '<v_nm>' ':' '<type>' [ '=' '<expr>' ] $var )# .
E : 'EVENT' ( '<e_nm>' $event '(' [ ( '<tag>' ':' '<type>'
    $attr1 )#( ',' $comma1 ) ] ')' $comma2 ) # .
A : 'ACTION' ( '<a_nm>' $action1 '(' [ ( '<tag>' ':' '<type>'
    $attr1 )#( ',' $comma1 ) ] ')' $comma2 ) # .
R : 'REQUEST' ( '<r_nm>' '(' ')' ':' '<type>' $request ) # .
Q : 'QUERY' ( '<q_nm>' $query '(' ')' ':' '<v_nm>' ) # .
U : 'STATE' ( ( '<u_nm>' $state1 ( '->' ( ( ( '<e_nm>' $u_nm1
    '(' [ ( '<val>' $attr2 )#( ',' $comma1 ) ] ')' $comma2 )# )
    [ '[' B ']' ] ; ( 'after' '(' '<val>' ( 's' ; 'ms' ) ')'
    $after ) ) [ Action ] $action4 '->' S ) ) )# ) $u_nm2 .
W : 'WAGGLY' ( 'entry' '->' [ '<e_nm>' $w_nm '(' [ ( '<val>'
    $attr2 )#( ',' $comma1 ) ] ')' $comma2 ] [ Action ] $action4
    '->' '<u_nm>' $start ) #( '<w_nm>' $state2 '->'
    [ 'after' '(' '<val>' ('s' ; 'ms') ')' $after ] [ Action ]
    $action4 '->' S $state4 ) .
D : 'DECISION' ( '<d_nm>' $d_nm1 '->' '(' ( '[' B ']' [ Action ]
    $action4 '->' S $state5 )#( '|' $else ) ( '|' $else
    [ '[' B ']' ] [ Action ] $action4 '->' S $state5 ) ')' )# .
G : 'GROUP' ( '<g_nm>' $g_nm '->' [ Action ] $action4
    '->' S $state6 )#( $d_nm2 ) .
L : 'LINK' ( ( '<f_num>' '<p_num>' ( '<e_nm>' ; '<q_nm>' )
    $link1 ) ; #( '<p_num>' '<t_num>' ( '<a_nm>' ; '<r_nm>' )
    $link2 ) )#( $close ).
B : $cond1 ( ( [ '!' $sign1 ] ( ( '<v_nm>' ( '<' ; '<=' ; '>' ;
    '>=' ; '==' ; '!=' ) $sign2 '<expr>' $cond2 ) ; ( '<r_nm>'
    '(' ')' $r_nm ) ) )#( '&&' $sign3 [ '!' $sign1 ] ) )#( '||'
    $sign4 [ '!' $sign1 ] ) .
Action : '/' ( ( '<v_nm>' '=' '<expr>' $action2 ) ; ( '<a_nm>'
    $action3 '(' [ ( ( '<v_nm>' ; '<val>' ; '<tag>' )
    $attr3 )#( ',' $comma1 ) ] ')' $comma2 ) )#( ',' ) .
S : ( '<u_nm>' ; '<w_nm>' ; '<d_nm>' ; '<g_nm>' ; 'exit') $state3 .
```

Для проверки правильности записи регулярных выражений использовалось инструментальное средство SynGT (Syntax Graph Transformations) [7,8]. На рисунке 1 показано графическое представление грамматики языка CIAO.



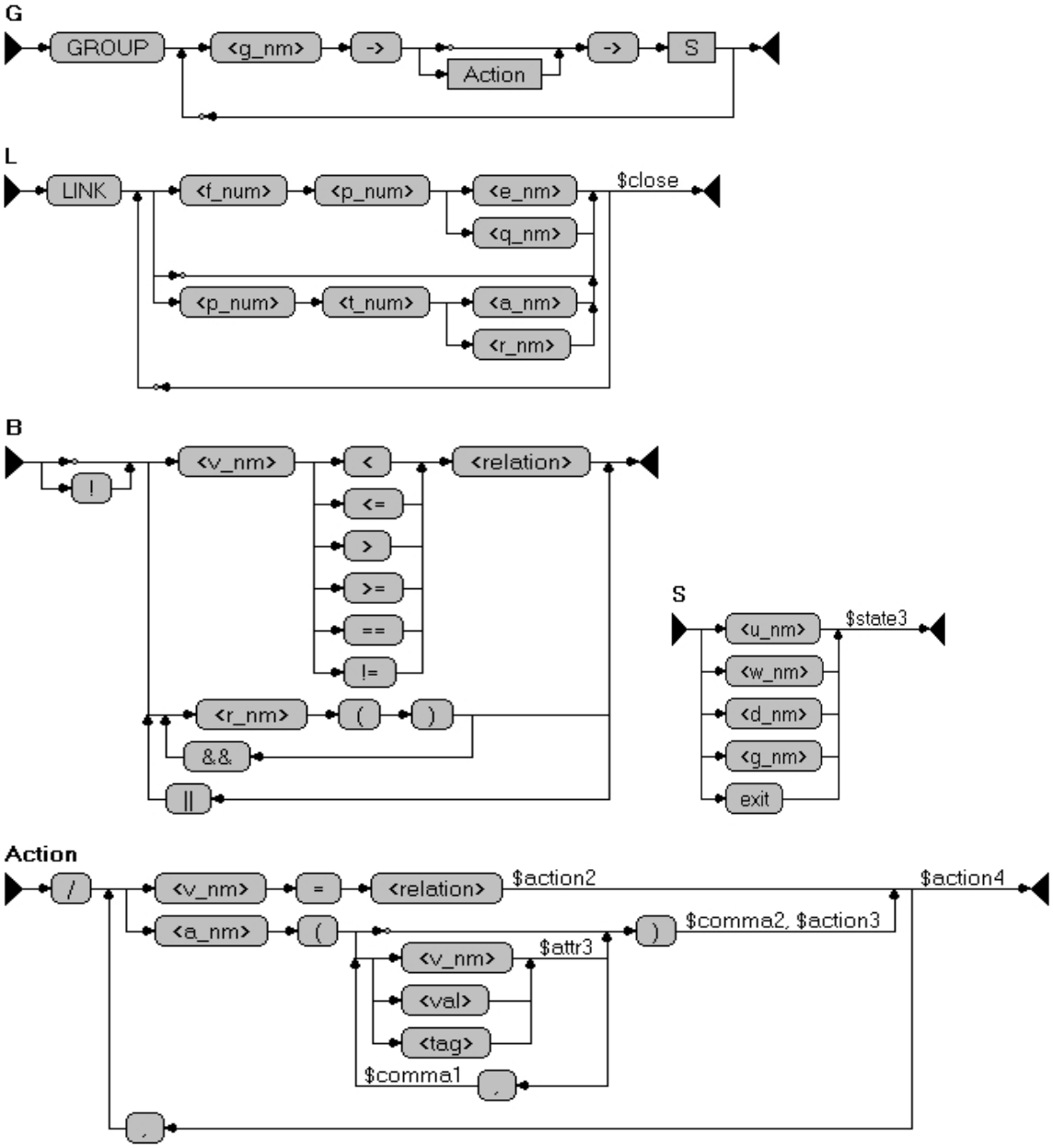


Рис. 1. Синтаксические графы для нетерминалов грамматики

Эксперименты по применению языка CIAO

В качестве примера выбрана система управления работой лифта, которая является достаточно показательной. Рассмотрим текстовую постановку задачи в несколько упрощённой форме, по сравнению с фундаментальной книгой [10].

Простой лифт в многоэтажном доме должен перевозить пассажиров с этажа на этаж. Лифт имеет кнопки вызова на каждом этаже и кнопки отправления на этажи в кабине. Система управления лифтом не принимает новых вызовов, пока не закончит обслуживание принятого вызова. Двери шахты и двери лифта открываются и закрываются системой управления, причём падение пассажиров в шахту лифта через несвоевременно открытые двери шахты или лифта должно быть исключено. В пол лифта встроен датчик давления. Система управления должна исключить движение с перегрузкой или недогрузкой лифта. В кабине лифта имеется освещение. Система управления должна включать и выключать освещение так, чтобы экономить электричество, но не пугать пассажиров несвоевременным выключением света.

Вначале по словесному описанию строится граф переходов состояний, отражающий поведение лифта (Рисунок 2). Методика явного выделения состояний для краткости опускается, в основных чертах она следует парадигме автоматного программирования [11].

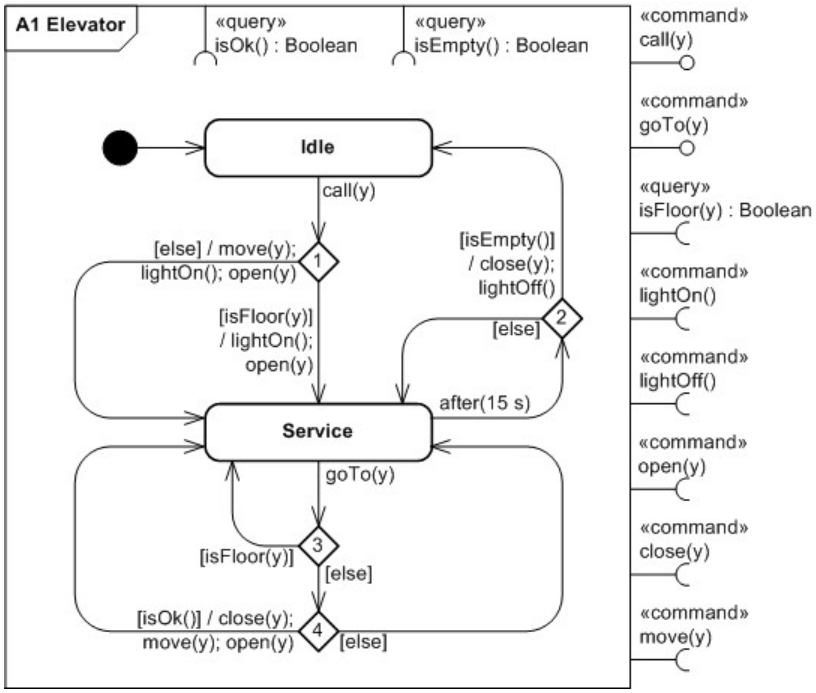


Рис. 2. Диаграмма состояний автоматного объекта Elevator

Графической диаграмме переходов состояний соответствует текстовое представление на языке CIAO (Листинг 1).

```
#1 Elevator
EVENT
call(y : Integer)
goTo(y : Integer)
ACTION
lightOn()
lightOff()
open(y : Integer)
close(y : Integer)
move(y : Integer)
REQUEST
isFloor(y) : Boolean
isOk() : Boolean
isEmpty() : Boolean
STATE
idle -> call(y) -> dcsn1
service -> after(15 s) -> dcsn2
service -> goTo(y) -> dcsn3
WAGGLY
entry -> -> idle
DECISION
dcsn1 -> ( [isFloor(y)] /
   lightOn(), open(y) -> service
   | / move(y), lightOn(), open(y
   ) -> service )
dcsn2 -> ( [isEmpty()] / close(y
   ), lightOff() -> idle | ->
   service )
dcsn3 -> ( [isFloor(y)] ->
   service | -> dcsn4 )
dcsn4 -> ( [isOk()] / close(y),
   move(y), open(y) -> service |
   -> service )
LINK
#2 #1 call
#2 #1 goTo
#1 #2 isFloor
#1 #2 lightOn
#1 #2 lightOff
#1 #2 open
#1 #2 close
#1 #2 move
```

Листинг 1 – Текстовое представление поведения автоматного объекта Elevator на языке CIAO

После применения кодогенератора SynGT к тексту в Листинге 1, получена заготовка текста для языка C++ (Листинг 2). Перед компиляцией в инструментальной среде следует вручную переопределить примитивы взаимодействия и синхронизации (многопоточность, механизмы обмена данными, таймеры), соответствующие используемой операционной системе.

```
typedef enum {
  idle, service
} State1;

class I1
{
public:
  void call(int y);
  void goTo(int y);
};

class A1 : public I1
{
private:
  State1 y_;
  I2 *out_;
  bool isOk();
  bool isEmpty();

public:
  void run() {
    y_=idle;
    int y;
    while (1) {
      event = wait(1);
      switch (y_) {
        case idle:
        {
          if (event==call(y)) {
            if (isFloor(y)) {
              out_->lightOn();
              out_->open(y);
              y_=service;
            }
            else {
              out_->move(y);
              out_->lightOn();
              out_->open(y);
              y_=service;
            }
          }
          break;
        }
        case service:
        {
          if (event==goTo(y)) {
            if (isFloor(y)) {
              y_=service;
            }
            else {
              if (isOk()) {
                out_->close(y);
                out_->move(y);
                out_->open(y);
                y_=service;
              }
              else {
                y_=service;
              }
            }
          }
          if (event==timer
             (15000)) {
            if (isEmpty()) {
              out_->close(y);
              out_->lightOff();
              y_=idle;
            }
            else {
              y_=service;
            }
          }
          break;
        }
      }
    }
  }
};
```

Листинг 2 – Программа на языке C++ для автоматного объекта Elevator

Замечательным свойством данной методики является то обстоятельство, что полученное решение является доказательно правильным (до некоторой степени). Действительно, требование безопасности “исключить падение пассажиров в шахту через несвоевременно открытые двери” выполнено, поскольку из диаграммы на рисунке 2 видно, что действие open(y) выполняется в любом случае только при выполнении предусловия isFloor(y). Все дальнейшие преобразования сохраняют это и другие свойства диаграммы переходов состояний.

Заключение

Описаны структура и текущее состояние грамматики языка CIAO для компиляции и проведения экспериментов с программами на языке CIAO для моделирования поведения сетей взаимодействующих автоматных объектов. Особенностью данной разработки является использование языка C++ с набором инструментальных средств для создания и отладки всех компонентов данной программы.

Дальнейшая работа будет состоять в расширении функциональности созданных компонентов (прежде всего кодогенератора и программы имитационного моделирования сгенерированного кода), уточнение грамматики и семантики входного языка CIAO и проведение экспериментов по программированию и исполнению реальных приложений.

Построенная автоматная модель таких расширенных диаграмм реализована в языке CIAO и обеспечивает модульность описания поведения программной системы, а также задаёт интерфейс взаимодействия как автоматов, так и любых других программных компонентов. Язык CIAO превосходит известные аналоги в части предоставления механизмов кооперативного взаимодействия, асинхронности и параллельности для широкого класса реагирующих систем.

ЛИТЕРАТУРА

1. Fedor A. Novikov, Ludmila N. Fedorchenko, Vladimir I. Vorobiev, Roza R. Fatkieva, and Dmitriy K. Levonevskiy. 2017. Attribute-Based Approach of Defining the Secure Behavior of Automata Objects. In Proceedings of SIN 2017 conference (SIN 2017), Jennifer B. Sartor, Theo D'Hondt, and Wolfgang De Meuter (Eds.). ACM, New York, NY, USA, Article 4, 6 pages. DOI: https://doi.org/10.1145/3136825.3136887
2. Афанасьева И.В., Новиков Ф.А. Архитектура программного обеспечения систем оптической регистрации // Информационно-управляющие системы. 2016. №3. С51–63. DOI: 10.15217/issue 1684-8853.2016.3.51
3. Новиков Ф.А., Афанасьева И.В. Кооперативное взаимодействие автоматных объектов // Информационно-управляющие системы. 2016. №6. С50–63. DOI: 10.15217/issn1684-8853.2016.6.50
4. Harel D. Statecharts: a Visual Formalism for Complex Systems // Science of Computer Programming. 1987. Vol. 8. P. 231–274.
5. Шалыто А. А. Парадигма автоматного программирования // Научно-технический вестник СПбГУ ИТМО. 2008. Вып. 53. С. 3-24.
6. Atiskov A. Y. et al. Ontology-Based Analysis of Cryptography Standards and Possibilities of Their Harmonization / A. Y. Atiskov, F. A. Novikov, L. N. Fedorchenko, V. I. Vorobiev, N. A. Moldovyan // Theory and Practice of Cryptography Solutions for Secure Information Systems. – Hershey: IGI Global, 2013.
7. Fedorchenko, L.: Regularization of Context-Free Grammars. LAP LAMBERT Academic Publishing, Saarbrucken. (2011).
8. Ludmila Fedorchenko and Sergey Baranov Equivalent Transformations and Regularization in Context–Free Grammars, // Bulgarian Academy of Sciences/ Cybernetics and Information Technologies (CIT) Volume 14, No 4, pp.11–28. Sofia 2015.
9. A. Aho, R. Sethi, J. Ullman, Compilers: Principles, Techniques and Tools. Addison-Wesley (1986).
10. Кнут Д. Искусство программирования для ЭВМ. Т.1 Основные алгоритмы. Перевод с английского. М. Мир, 1976.
11. Поликарпова Н. И., Шалыто А. А. Автоматное программирование. – СПб.: Питер, 2011. – 176 с.

I.V.Afanacieva (SAO RAS, Nizhniy Arkhiz, ITMO University), **F.A. Novikov** (St. Petersburg Polytechnic university of Peter the Great), **D.K. Levonevskiy, L.N. Fedorchenko** (SPIIRAS)

Formal model and methods for the system description with complex behavior on principles of generalized automata.

Automated models are most convenient for designing systems with complex behavior, but have a number of limitations: the lack of support for asynchrony and parallelism. The proposed report is devoted to the description of the CIAO high-level programming language, from the program of which a model of the system of automata simulating the functioning of the initial reacting system is generated

АЛГОРИТМ ПОСТРОЕНИЯ НЕОДНОРОДНОЙ МАТЕМАТИЧЕСКОЙ СТРУКТУРЫ АЛГЕБРАИЧЕСКИХ БАЙЕСОВСКИХ СЕТЕЙ

Н. А. Харитонов[1]
(*СПИИРАН, Санкт-Петербург, 14-я линия В.О., 39 nikita.kharitonov95@yadex.ru*)

Аннотация

В данной работе представлены алгоритмы построения алгебраических байесовских сетей с цепной и звездчатой структурой. Даны определения данных структур, приведены результаты замеров времени работы алгоритмов в зависимости от количества атомов в алгебраической байесовской сети и во фрагменте знаний.

Ключевые слова: вероятностные графические модели, алгебраические байесовские сети, автоматизация.

Введение

Алгебраические байесовские сети[1] являются вероятностными графическими моделями, представленными ненаправленными графами с идеалами конъюнктов в узлах, которым приписываются скалярные или интервальные оценки вероятностей. В целях оптимизации сеть делится на малые части, называемые фрагментами знаний, и представляющие из себя полный идеал конъюнктов над заданным набором атомов. Локальный вывод[2,3] – это процессы, происходящие во фрагментах знаний, глобальный – процессы, происходящие во всей сети. При исследовании последних (процессов поддержания непротиворечивости, априорного и апостериорного вывода) возникает необходимость создания примеров алгебраической байесовской сети. Создание примеров вручную является трудоемким процессом, в связи с чем было проведено данное исследование, целью которого является автоматизация создания алгебраической байесовской сети, построенной над фрагментами знаний различного размера.

Цепная и звездчатая структуры.

В исследовании рассматривалось два вида структуры алгебраической байесовской сети: цепная и звёздчатая.

АБС с цепной структурой представляет собой последовательно идущие фрагменты знаний, каждый из которых имеет пересечение только с двумя другими, «соседними». При этом два фрагмента знаний являются «крайними» и имеют единственное пересечение.

АБС со звёздчатой структурой представляет собой набор фрагментов знаний, каждый из которых связан со всеми остальными. При этом пересечение является единственным и общим для всех фрагментов знаний.

В рамках данной работы количество атомов на пересечениях для АБС с цепной структурой является одинаковым.

Алгоритмы построения

Процессы построения АБС с цепной и со звездчатой структурами схожи между собой, в данном разделе будет описан алгоритм создания АБС с цепной структурой.

Входными данными для алгоритма являются число атомов в АБС (n), число атомов в ФЗ (m) и число атомов на пересечениях(k), причем n-(m-k) должно быть кратным m.

Первым шагом алгоритма является создание «пустой» алгебраической байесовской сети над n атомами.

Далее в АБС добавляются генерируемые автоматически фрагменты знаний. Все создаваемые ФЗ имеют индекс, состоящий из (n-m) нулей и m единиц. Каждое число индекса соответствует атому из АБС, равно единице в случае, если этот атом присутствует во фрагменте знаний, и нулю в противном случае.

[1] Младший научный сотрудник

Единицы в индексе идут подряд и для первого фрагмента знаний занимают первые m позиций индекса. Это позволяет путем смещения единиц в индексе у следующего фрагмента знаний на (m-k) позиций вправо создавать цепную структуру, поскольку в таком случае k атомов будут принадлежать обоим фрагментам знаний.

Здесь стоит отметить, что для звездчатой структуры k первых позиций индекса всегда будут заняты единицами, что является пересечением, а оставшиеся (m-k) единиц будут для каждого последующего ФЗ смещаться вправо на (m-k).

В каждом создаваемом фрагменте знаний верхняя и нижняя оценки интервала истинности конъюнктов генерируются случайным образом, при этом верхняя граница не может быть меньше нижней.

Время работы алгоритма

В данном разделе представлены замеры времени работы алгоритма в миллисекундах.

Генерация алгебраической байесовской сети проводилась в зависимости от числа атомов во фрагменте знаний. Количество атомов на пересечениях считалось равным единице.

Замеры времени производились 50 раз. Далее на основе полученных результатов вычислялись математическое ожидание, среднее квадратичное отклонение и дисперсия. Полученные величины продемонстрировали состоятельность результатов, в таблицах и на графиках ниже представлены полученные математические ожидания.

Таблица 1

Результаты работы алгоритма для АБС с цепной структурой

Число атомов в АБС	2 атома в ФЗ	3 атома в ФЗ	4 а тома в ФЗ
4	403	355	461
5	492	471	634
6	594	581	808
7	684	754	978
8	791	843	1272
9	940	944	1548
10	1012	1041	1870

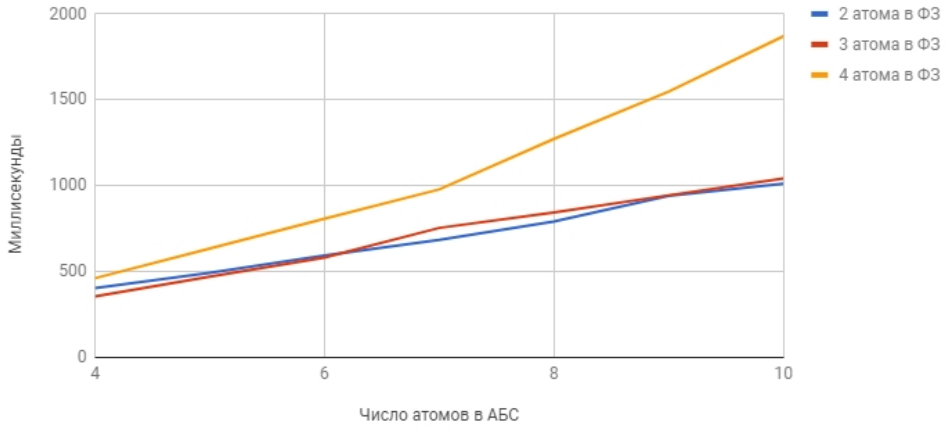


Рис. 1. Результаты работы алгоритма для АБС с цепной структурой, график

Таблица 2

Результаты работы алгоритма для АБС со звездчатой структурой

Число атомов в АБС	2 атома в ФЗ	3 атома в ФЗ	4 а тома в ФЗ
4	370	363	511
5	538	478	677
6	618	591	856
7	761	705	1027
8	831	821	1191
9	937	933	1341
10	1052	1042	1523

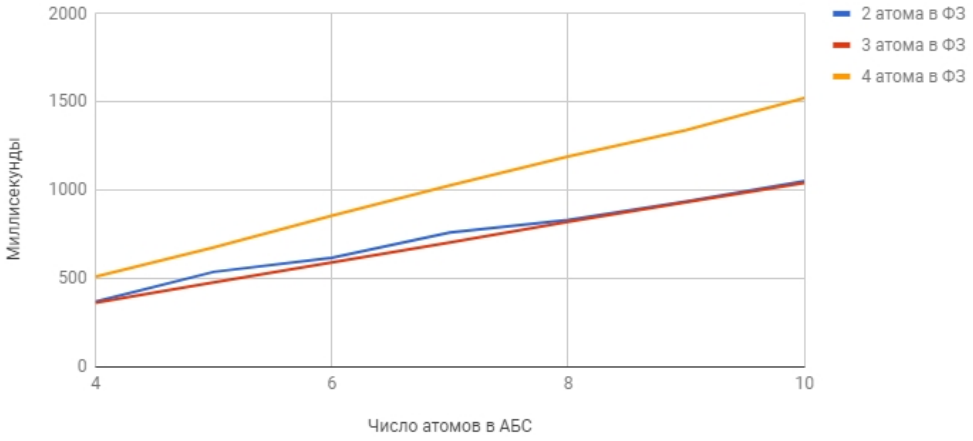


Рис. 2. Результаты работы алгоритма для АБС со звездчатой структурой, график

Заключение

Результаты проведенного исследования расширяют возможности по исследованию непротиворечивости, априорного и апостериорного вывода в алгебраических байесовских сетях. В дальнейшем планируется создать алгоритм, создающий абсолютно случайную сеть.

Сами алгебраические сети в дальнейшем предполагается использовать, например, при моделировании социоинженерных атак[4].

Часть результатов получена при поддержке гранта РФФИ 18-01-00626 и в рамках проекта по государственному заданию СПИИРАН № 0073-2018-0001.

ЛИТЕРАТУРА

1. Тулупьев А.Л., Николенко С.И., Сироткин А.В. Байесовские сети: логико-вероятностный подход. СПб: Наука, 2006. 607 с.
2. Мальчевская Е.А., Золотин А.А. Уравнения апостериорного вывода в фрагментах знаний над идеалом дизъюнктов // Материалы Всероссийской научной конференции по проблемам информатики СПИСОК. 2017. С. 395-403.
3. Сироткин А.В., Тулупьев А.Л. Локальный априорный вывод в расширенном фрагменте знаний с вероятностной неопределенностью // Сборник научных трудов сессии МИФИ в 15 томах «Интеллектуальные системы и технологии». 2008. Т. 10. С. 144-145.
4. Abramov M. V., Azarov A. A. Social engineering attack modeling with the use of Bayesian networks //Soft Computing and Measurements (SCM), 2016 XIX IEEE International Conference on. – IEEE, 2016. – С. 58-60..

N.A.Kharitonov (SPIIRAS, Saint-Petersburg)

Algorithm for constructing an heterogeneous mathematical structure of algebraic Bayesian networks

In this paper, algorithms for constructing algebraic Bayesian networks with a chain and stellate structure are presented. The the results of measurements of the operation time of the algorithms are given depending on the number of atoms in the algebraic Bayesian network and in the knowledge pattern.

Секция 2
СИСТЕМНЫЙ АНАЛИЗ И КОМПЛЕКСНОЕ МОДЕЛИРОВАНИЕ АВТОМАТИЗИРОВАННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ЛОГИСТИЧЕСКИХ СИСТЕМАХ

В. С. Лукинский[1]
(*Национальный исследовательский университет Высшая школа экономики,
194100, С.-Петербург, ул. Кантемировская, д.3, к.1, лит.А
vslukinskiy@hse.ru*)

Ю. М. Искандеров[2], **Б. В. Соколов**[3]
(*Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук,
199178, С.-Петербург, 14 линия, д. 39
iskanderov_y_m@mail.ru, sokolov_boris@mail.ru*)

А. Г. Некрасов[4]
(*Московский автомобильно-дорожный государственный технический университет «МАДИ»,
125319, г. Москва, Ленинградский пр., 64
tehnologistic@mail.ru*)

Аннотация

В докладе предлагается подход, направленный на разработку и исследование научно-методического обеспечения процессов проактивного интеллектуального управления логистическими функциями в цепях поставок на основе внедрения и использования релевантных интеллектуальных систем, позволяющих интегрировать (объединить) фундаментальные и прикладные результаты, полученные в настоящее время в логистическом менеджменте и теории управления сложными системами. Дана характеристика современного состояния решаемой проблемы и изложено обоснование выбранного подхода. В докладе представлены результаты, которые позволяют формировать научно-методическое обеспечение использования интеллектуальных информационных технологий в логистических системах с целью их эффективного управления, а также перспективы развития исследований, обеспечивающих создание новых технологий социально-экономического развития в рассматриваемой предметной области.

Ключевые слова: интеллектуальные информационные технологии, логистические системы.

Введение

За последние несколько лет достигнут значительный прогресс в развитии и теоретической и методологической составляющих логистического менеджмента. В многочисленных трудах отечественных и зарубежных ученых, в том числе и [1-13], подробно исследованы вопросы генезиса и эволюции развития логистики в целом, а также отдельных ее функциональных областей, моделей и методов; сформирован понятийный аппарат теории; разработаны классификации и описаны свойства основных потоковых процессов; исследованы подходы к формированию научной базы методологии логистики. Между тем, анализ современного состояния исследований в логистике показывает:

1. Недостаточно высокий уровень прикладных моделей и методов, используемых в логистике. К тому же большинство аналитических зависимостей в иностранных работах по логистике – это модели и методы управления запасами, в то же время другим логистическим функциям

[1] Доктор технических наук, профессор, руководитель департамента.
[2] Доктор технических наук, профессор, заведующий лабораторией.
[3] Доктор технических наук, профессор, заведующий лабораторией.
[4] Доктор экономических наук, профессор, профессор кафедры.

(транспортировка, закупка, складирование и др.) уделяется существенно меньшее внимание; представленный аналитический аппарат характеризуется разрозненностью и незавершенностью.

2. Известно, что новизна логистического подхода заключается в интеграции закупочной (снабженческой), производственной, распределительной (сбытовой) деятельности. Между тем, до последнего времени методология логистики представляет собой скорее набор отдельных моделей (методов, алгоритмов), практически не систематизированных и недостаточно подробно изученных. При этом не существует ни одной пространственно-временной модели, описывающей поведение материального (и сопутствующих) потока и охватывающей несколько звеньев логистической системы, без чего трудно добиться прогресса при решении основной проблемы - оптимизации издержек в цепях поставок.

3. В настоящее время за рубежом в рассматриваемой предметной области применительно, например, к наиболее сложным задачам конфигурирования виртуальных предприятий реализуется ряд концепций (Wireless Enterprises, M2M, Mobile Commerce, Virtual Information Technologies, Virtual Enterprise), в рамках которых предусматривается обеспечение коммуникации и интеграции, совместное управление, моделирование взаимодействия участников сетевых (виртуальных) логистических систем на основе использования Internet-технологий. Все они ориентированы на решение задач формирования и организации функционирования рассматриваемых предприятий, основанных на использовании единой организационно-технологической и информационной среды, за счет временного объединения ресурсов различных предприятий. Однако, существующие концепции логистического менеджмента предусматривают решение рассматриваемых проблем с разных точек зрения – от простого согласования планов сбыта и поставок несколькими компаниями до всеобъемлющей концепции расширяющегося логистического взаимодействия (SLI – Scaleable Logistics Interplay), которую планируется реализовать в 21 веке. В рамках этих концепций ведутся интенсивные исследования и разработки по всему спектру сформулированных в данном проекте задач. В частности, текущие исследования в области информационно-аналитического обеспечения виртуальных предприятий (ВП) сфокусированы на следующих аспектах:

• теоретические и практические основы (разработка моделей предприятия и протекающих в нем процессов, а главное – моделей взаимодействия предприятий);

• обеспечение совместного выполнения бизнес-процессов;

• обеспечение информационной безопасности бизнес-процессов;

• обеспечение процессов управления комплексными производственно-логистическими системами.

4. Процесс создания комплексной информационной сети с участием поставщиков, производителей, торговых организаций и клиентов является чрезвычайно сложным организационно-технологическим процессом и требует тщательной предварительной подготовки, в первую очередь, разработки соответствующего научно-методического обеспечения.

С технической точки зрения, создание единого интеллектуального информационного пространства, обеспечивающего успешную реализацию концепции расширяющегося логистического взаимодействия, должно, в первую очередь, основываться, на разработке инструментальных средств обмена разноформатными данными и знаниями гетерогенных логистических информационных систем, базирующихся на аппаратно-программных архитектурах различных производителей. В частности, для этого используется ряд известных информационных технологий, таких как CORBA, J2EE (Sun), Dot.Net (Microsoft) и др., а также специальные форматы данных, например, XML. Однако, указанные технологии не охватывают в значительной степени глубину процессов принятия решений в логистическом менеджменте, и к сожалению, готовых методологических, методических способов и инструментальных средств, являющихся объектом исследования в данном проекте, ни за рубежом, ни в нашей стране пока нет.

Более того постановка проблемы создания научно-методического обеспечения процессов проактивного управления логистическими функциями в цепях поставок с использованием интеллектуальных систем даже не осуществлялась.

Необходимость в разработке такого научно-методического обеспечения обусловлена:

- недостаточной проработкой (в отдельных случаях – отсутствием) методов и принципов управления логистическими функциями в цепях поставок с использованием интеллектуальных систем для решения актуальных и перспективных задач логистического менеджмента;

- нескоординированным развитием и использованием существующих разнотипных автоматизированных систем управления различных поколений на уровне отдельных субъектов хозяйствования, что может привести в будущем без соответствующей координационной политики к дополнительным проблемам и издержкам при их объединении не только в рамках единого корпоративного информационного пространства, но и при обычном информационном взаимодействии отдельных элементов логистических цепей.

Таким образом, настоящий доклад посвящен проблемам разработки научно-методических основ обеспечения процессов управления логистическими функциями в цепях поставок с использованием интеллектуальных информационных технологий на основе реализации концепций системного моделирования и интеграции знаний применительно к рассматриваемой предметной области.

Характеристика области исследования

Основным объектом исследования выбраны существующие и перспективные логистические функции (формирование запасов, транспортировка, закупка, складирование и др.) в цепях поставок, представляющие собой сложные взаимосвязанные организационно-технические сетевые структуры, включающие в себя элементы различной природы и назначения, в том числе материальные объекты, информационно-управляющие системы, , а также социальные (организационные) структуры, которые определяют целевое предназначение указанных систем и являются основными потребителями результатов их деятельности (потребителями товаров и услуг).

Предметом исследований в проекте определены методические и технологические основы внедрения и использования средств интеллектуальных систем в процессы проактивного управления логистическими функциями в цепи поставок в интересах совершенствования технологий логистического менеджмента при переходе к цифровой экономике.

Для достижения поставленной цели необходимо решить следующие основные задачи:

• выполнить системный анализ современных подходов к решению проблем анализа и синтеза (конфигурирования) систем проактивного управления логистическим функциями в цепях поставок с использованием интеллектуальных систем (ИС) в различных предметных областях;

• разработать требования к облику перспективных систем проактивного управления логистическим функциями в цепях поставок;

• разработать концептуальную модель системы проактивного управления, ориентированной на оптимизацию производственно-транспортных (логистических) процессов в цепях поставок;

• разработать и обосновать методологические основы (принципы, концепции, способы, подходы) использования и внедрения ИС по проактивному управлению логистическими функциями в цепях поставок;

• разработать и обосновать критерии и показатели оценки эффективности интеграции технологий логистического менеджмента и методов проактивного управления;

• обосновать состав и структуру модельно-алгоритмического и информационного обеспечения системы проактивного управления логистическим функциями в цепях поставок с использованием ИС;

• разработать прототип системы проактивного управления логистическим функциями в цепях поставок с использованием ИС;

• разработать предложения по внедрению и использованию системы проактивного управления логистическим функциями в цепях поставок с использованием ИС.

Исходя из этого, в целом проект направлен на решение актуальной научной проблемы интеллектуализации подготовки и принятия решений в системах социально-экономического управления и прогнозирования их развития с учетом формирования цифровой экономики.

В настоящее время существует множество частных подходов к формализации и исследованию проблем анализа и синтеза (проблем конфигурирования) интеллектуальных информаци-

онных технологий (ИИТ) в логистических систем (ЛС). Данные подходы, как правило, базируются на количественных моделях, описывающих рассматриваемую проблемную область. К указанным моделям, в первую очередь, могут быть отнесены модели, используемые в исследовании операций, теории управления запасами, математической статистики. Однако, как показывает анализ, решение реальных задач конфигурирования ИИТ ЛС, возникающих на практике, базируется в большей степени на разнородных и распределенных декларативных и процедуральных знаниях, и главная проблема, связанная со структурно-функциональным синтезом облика ИИТ ЛС, состоит как раз в интеграции указанных знаний как на уровне бизнес-процессов, так и на уровнях соответствующих приложений и аппаратно-программных платформ с использованием современных интеллектуальных информационных технологий и инструментальных средств.

С учетом выше изложенного в настоящее время особую актуальность приобретают вопросы разработки методологических и методических основ обоснования и выбора наиболее перспективных путей создания и развития ИИТ ЛС на основе реализации концепций системного моделирования и интеграции знаний применительно к рассматриваемой предметной области.

В СПИИРАН в настоящее время проводятся комплексные исследования в области транспортно-логистических процессов и систем на базе лабораторий: “Интегрированных систем автоматизации”, “Интеллектуальных систем”, “Информационных технологий на транспорте”, “Проблем компьютерной безопасности”, “Информационных технологий в системном анализе и моделировании”, “Прикладной информатики”. В перечисленных лабораториях в настоящее время проводятся интенсивные исследования, направленные на создание и совершенствование следующих современных информационных технологий (ИТ), широко используемых в различных предметных областях (в том числе и в области логистики): ИТ извлечения знаний из данных, машинного обучения, ИТ, базирующиеся на многоагентных системах, компьютерное зрение, речевая информатика, стеганография и стеганоанализ, интеллектуальные сенсорные сети, ИТ защиты компьютерных сетей, новые технологии компьютерного моделирования и супервычислений, биометрия и т.д.

В организации перечисленных работ СПИИРАН активно взаимодействует с другими организациями, среди которых, в первую очередь, можно выделить Высшую школу экономики, а также Московский автодорожный институт. В рамках данной кооперации удалось к настоящему времени сформулировать, обосновать и приступить к выполнению ряда проектов, направленных на повышение эффективности транспортно-логистических процессов и систем в РФ и Северо-Западном регионе.

Предлагаемые подходы и методы, их обоснование

Выполненные исследования показали, что возможны различные (гипотетические) варианты решения задач проекта. Первый вариант - эволюционный, который будет характеризоваться преобладающим развитием информационных технологий, с полной автоматизацией процедур принятия и поддержки решений. Второй вариант, наиболее предпочтительный с точки зрения авторского коллектива, – это качественный скачок в изучении и развитии прикладных моделей и методов логистического менеджмента, которые не получили должного развития с 60-х годов 20-го века с последующим синтезом современного инструментария на основе новых информационных технологий, (первая волна: в 20-х годах 20 века – труды таких ученых, как Харрис, Уилсон и др., вторая волна – 60–е годы 20 века: труды, посвященные управлению запасами).

Интенсификация развития процесса формирования научного знания логистического менеджмента, а также механизма его трансформации в практический бизнес-инструментарий управления потоковыми процессами в реально функционирующих социально-экономических системах требует формализации базовой логистической концепции общих логистических затрат, реализация которой предполагает:

- исследование структурной взаимосвязи модели общих логистических затрат и модели Харриса-Уилсона;
- разработку структуры обобщенной модели логистических издержек, ядром которой должна являться модель оптимального размера заказа (EOQ), учитывающей уровень логистического сер-

виса (обслуживания) и временные параметры выполнения логистических циклов (на основе концепций «точно вовремя» (JIT), «от двери до двери» (DTD), «быстрого реагирования» (QR) и др.);

- пересмотр функциональных областей логистики: снабжения, поддержка производства, распределения; ключевых и поддерживающих логистических функций (складирования, транспортировки и др.) с целью формирования нового аналитического аппарата (современных моделей и методов), отражающих интегральный характер цепей поставок;

- формирование комплексов аналитических моделей управления отдельными логистическими функциями в цепях поставок: складированием, транспортировкой и др.

Анализ показывает, что эффективная реализация систем управления конфигурацией логистических сетей возможна только на основе интеграции современных информационных технологий и их дальнейшей интеллектуализации. Такими перспективными технологиями являются: технология совмещенного проектирования (concurrent engineering, hardware and software co-design), технология удовлетворения ограничений (constraint satisfaction), технологии системного моделирования и интеллектуального управления, технология создания объектно-ориентированных и интеллектуальных баз данных, технология интеллектуальных геоинформационных систем, технология проектирования и применения многоагентных и гибридных систем.

В основу интеллектуальной информационной технологии должны быть положены подходы проактивного управления и интегрированной логистики, оцениваемые в настоящее время как наиболее перспективные. Они развиваются в основном в рамках решения проблем искусственного интеллекта и его практических приложений в разделах, ориентированных на исследование задач кооперативного принятия решений в распределенной среде с использованием "интеллектуальных" и "мобильных агентов", а также многоагентных, гибридных и геоинформационных систем и технологий. С данным подходом связаны большие ожидания в области искусственного интеллекта и телекоммуникаций, поскольку предлагаемая парадигма рассматривается как наиболее перспективный путь слияния достижений в отмеченных выше направлениях, как комплексный подход, который приведет к появлению качественно новых возможностей в области информатики и современной кибернетики, тем самым способствуя развитию цифровой экономики.

Весьма перспективной также является и технология системного моделирования сложных социально-экономических организаций, позволяющая на конструктивном уровне проводить их полимодельное описание и исследование, как с использованием традиционных математических моделей, так и моделей, базирующихся на “мягких вычислениях” (soft computing). Данную технологию предлагается использовать при разработке методологических основ решения проблем конфигурирования структуры системы проактивного управления логистическими функциями в цепях поставок, а также при обосновании облика модельно-алгоритмического и информационного обеспечения указанной системы.

Кроме того, необходимо разработать новую концептуальную модель системы проактивного управления, ориентированной на оптимизацию производственно-транспортных (логистических) процессов в цепях поставок, которая будет базироваться на объектно-ориентированном подходе к описанию исследуемой предметной области. С использованием данной модели предлагается учитывать как иерархические, так и одноуровневые взаимосвязи ЛС, которые формально задаются с помощью комбинированных моделей знаний, а также осуществлять интеграцию различных подзадач, входящих в состав задачи конфигурирования. Данная модель базируется как на существующих инструментальных системах, поддерживающих технологии удовлетворения - распространения ограничений, недоопределенных вычислений, мультиагентные технологии, используемые для поиска допустимых и эффективных решений в ходе конфигурирования структуры системы проактивного управления логистическими функциями в цепях поставок на различных этапах их жизненного цикла, так и на разработанных в рамках проекта онтологиях, описывающих различные аспекты данной проблемы**.**

Полученные результаты

За последние годы авторами, опираясь на накопленный потенциал, создан теоретический и прикладной задел для комплексного решения проблем в данном проекте. Авторы исследования

принимали участие как научные руководители и ответственные исполнители в различных проектах, в том числе, таких как: «Методология интеграции транспортных систем России с учетом развития глобального рынка транспортных услуг» по программе Президиума РАН «Фундаментальные проблемы пространственного развития РФ: междисциплинарный синтез».

В ходе указанных проектов получены следующие важнейшие результаты:

1) Принципы построения релевантных логистических моделей организации транспортных процессов, учитывающих разнородность элементов и связей между ними, распределенность в пространстве и во времени, иерархически-сетевой принцип организации, влияние природных и техногенных факторов. В качестве инструментального средства реализации моделей выбраны мультиагентные технологии, отличительной особенностью которых является многовариантность построения информационных систем и эффективное моделирование сетевых организаций. Новизна проведенных исследований состоит в том, что впервые мультиагентные технологии применены для разработки логистических моделей организации транспортных процессов. Практическая значимость определяется возможностью применения полученных результатов при моделировании и оценке эффективности функционирования сложных транспортных систем.

2) Принципы построения модели интегрированной информационной системы, обеспечивающей рациональное взаимодействие различных видов транспорта в едином информационном пространстве. Реализация принципов позволяет динамически перестраивать информационную систему при одновременной унификации программного обеспечения, управлять информационными ресурсами в интересах эффективного функционирования и развития транспорта и обеспечить вхождение в международное информационное пространство с целью развития мобильных услуг.

3) Модель синтеза интегрированной информационной системы транспортной логистики, использующая мультиагентные технологии и учитывающая основные ее характеристики: достоверность, скорость и своевременность доставки сообщений, надежность и устойчивость связи. Модель позволяет унифицировать используемое программное обеспечение, динамически перестраивать структуру ИС и встраивать в нее ранее разработанные средства и программы и, как следствие этого повысить эффективность функционирования и развития транспортной системы за счет оптимизации управления распределенными информационными ресурсами. Модель легла в основу оценки эффективности транспортно-технологических процессов в ОАО «РЖД».

4) Системно-структурная модель процедуры интеллектуальной поддержки принятия решений по организации транспортных процессов на основе принципов логистики «от двери до двери» и «точно в срок». Формализация указанной процедуры выполнена в виде общей задачи принятия решений. Выявлены и исследованы особенности функционирования разработанной модели, для представления знаний и манипулирования ими в модели использован оригинальный инструментарий на базе семантических графов с оболочками. Разработанная модель позволяет обеспечить синтез рациональных решений по организации транспортных процессов с учетом их целей, задач, функциональной структуры и пространственно-временных характеристик.

5) Системный анализ организации транспортных процессов при взаимодействии различных видов транспорта. На основе теории конфликтов и теории компромиссов установлены особенности такой организации и с применением методов нечетких множеств и нечеткой логики проведена формализация выявленных особенностей. Полученные результаты позволяют учесть и сбалансировать противоречивость интересов различных участников транспортного процесса и могут быть использованы при разработке моделей единого транспортного пространства страны.

Указанные научные результаты были внедрены в практическую деятельность ОАО "Российские железные дороги" и нашли непосредственное отражение в стратегических направлениях научно-технического развития ОАО «РЖД» на период до 2015 г. («Белая книга» ОАО «РЖД»). Стратегия предусматривает комплекс мер по совершенствованию перевозочного процесса, повышению надежности и увеличению ресурса технических средств, развитию корпоративной системы управления качеством, развитию высокоскоростного движения и др.

Полученные на сегодняшний день результаты подтверждают обоснованность предлагаемых подходов и методов исследований, и свидетельствуют о своевременности проведения исследо-

ваний по созданию целостного научно-методического обеспечения процессов принятия решений при проактивном управлении логистическими функциями в цепях поставок с использованием интеллектуальных систем.

Перспективы развития исследований

В соответствии с поставленной целью исследования и решаемыми задачами, при разработке научно-методического обеспечения процессов проактивного интеллектуального управления логистическими функциями в цепях поставок предполагается направить усилия на получение следующих научных результатов:

• на основе системного анализа современных подходов к решению проблем анализа и синтеза (конфигурирования) систем проактивного управления логистическим функциями в цепях поставок с использованием интеллектуальных систем в различных предметных областях раскрыть возможности предлагаемого подхода, выявить ключевые элементы логистического менеджмента, определяющие формирование новых технологий управления с использованием средств искусственного интеллекта, при этом показать изменение целей и способов функционирования сложных логистических структур, перераспределение функций, задач и алгоритмов управления между их элементами и др.;

• сформировать новые требования к облику перспективных систем проактивного управления логистическим функциями в цепях поставок, определяющие интеграцию и эффективное использование накопленных знаний в области логистического менеджмента и средств новых интеллектуальных технологий. Будет выполнена классификация указанных требований, позволяющая осуществлять адекватный выбор средств интеллектуальных систем в интересах повышения эффективности управления как отдельными логистическими функциями, так и процессом интегральной логистики в целом;

• разработать концептуальную модель системы проактивного управления, ориентированную на оптимизацию производственно-транспортных (логистических) процессов в цепях поставок в интересах снижения общих логистических затрат. В состав указанной модели будут включены модели, характеризующие использование элементов искусственного интеллекта, модели, реализующие технологии применения накопленного опыта в логистическом менеджменте, модели социально-экономического взаимодействия участников логистического процесса, модели взаимодействия «человек-техника» и др.;

• сформировать методологические основы (принципы, концепции, способы, подходы) использования и внедрения ИС по проактивному управлению логистическими функциями в цепях поставок, в том числе такие как принцип программно-целевого управления, внешнего дополнения, необходимого разнообразия, неокончательных решений, принцип полимодельности и многокритериальности, принцип новых задач, на базе которых могут быть разработаны новые эффективные инструментальные средства логистического менеджмента;

• обосновать критерии и показатели оценки эффективности интеграции технологий логистического менеджмента и методов проактивного управления, отличающиеся возможностью и способностью учета реализации внедрения и использования средств ИС. На основе теории конфликтов и теории компромиссов будут установлены особенности проактивного управления в указанных условиях, что позволит учесть и сбалансировать противоречивость интересов различных участников, реализующих логистические функции в цепях поставок;

• определить состав и структуру модельно-алгоритмического и информационного обеспечения системы проактивного управления логистическим функциями в цепях поставок, содержащие новые элементы средств ИС, обеспечивающие функционирование процессов управления операциями взаимодействия, ресурсами, потоками, структурной динамикой, движением, каналами, параметрами операций, вспомогательными операциями и др.;

• разработать прототип системы проактивного управления логистическим функциями в цепях поставок с использованием ИС, при создании которого должны быть использованы технологии распределенной разработки и обеспечено непосредственное участие экспертов (аналитиков) и инженеров по знаниям в концептуальном и логическом проектировании онтолого-ориентированных баз знаний, построении сценариев интеллектуальной оперативно-

аналитической обработки информации, и реализация которого должна быть выполнена на базе имитационно-модельного стенда, развернутом в региональном ситуационном центре;

• разработать предложения по внедрению и использованию системы проактивного управления логистическим функциями в цепях поставок с использованием ИС, направленные на генерацию новых методов и технологий логистического менеджмента, позволяющих эффективно использовать возможности и достоинства как средств искусственного интеллекта, так накопленный опыт в виде экспертных знаний специалистов в области интегрированной логистики.

Получение указанных результатов позволит реализовать формирование научно-методического обеспечения процессов системы проактивного управления логистическим функциями в цепях поставок с использованием ИС, обеспечивающего создание новых технологий социально-экономического развития в своей предметной области.

Заключение

Представленный подход направлен на разработку и исследование научно-методического обеспечения процессов проактивного интеллектуального управления логистическими функциями в цепях поставок (ЛФ ЦП) на основе внедрения и использования релевантных интеллектуальных систем, позволяющих интегрировать (объединить) фундаментальные и прикладные результаты, полученные в настоящее время в логистическом менеджменте и теории управления сложными системами.

Актуализация проблемы поиска реальных путей и разработки действенных механизмов повышения эффективности и конкурентоспособности отечественных предприятий вызвана интенсивным нарастанием глобализационных процессов в мировой экономике, усилением конкуренции на международных рынках, сопровождающимся возрастанием требований потребителей к уровню предоставляемого сервиса, а также необходимостью использования принципиально новых прогнозирующих и упреждающих управляющих воздействий, базирующихся на методологии и технологиях комплексного (системного) моделирования, включающих в себя полимодельное описание исследуемой предметной области, а также комбинированное использование методов, алгоритмов и методик многокритериального оценивания, анализа и выбора наиболее предпочтительных решений. В настоящее время информация и знания все в большей мере становятся стратегическим ресурсом общества, его движущей производительной силой. На смену индустриальному этапу развития общества пришла новая эволюционная фаза, фаза информатизации и соответствующая общественно-экономическая формация – информационное общество, при котором наиболее эффективное и динамичное его развитие возможно на основе максимально полного использования имеющихся информационных ресурсов и средств их обработки, составляющих основу соответствующих информационных пространств. Главным ресурсом ускоренного развития современного информационного общества становятся знания, главным механизмом развития – цифровая экономика, основанная на знаниях. Главными технологиями цифровой экономики становятся новые информационные и коммуникационные технологии, которые коренным образом меняют содержание и форму государственного регулирования экономики и управления социально-экономическими процессами, а также прогнозирование их развития, основанное на применении достижений в области систем искусственного интеллекта, позволит существенным образом повысить качество управления экономикой, в том числе логистическими функциями в цепях поставок. Одним из ключевых направлений логистического менеджмента является сокращение соответствующих издержек. По оценкам специалистов, логистические издержки в Российской Федерации составляют около 20 % от валового внутреннего продукта. При условии снижения уровня затрат на логистику до среднемирового уровня (т.е. на 6 %) экономия в 2017 г. могла бы составить более 15 трлн. руб., что сопоставимо с затратами на основные социальные программы: здравоохранение, образование и др. Снижение уровня логистических издержек может быть достигнуто активным внедрением и использованием в процессы функционирования современных предприятий интеллектуальных систем. Принципиальными особенностями изложенного подхода, определяющими его новизну, являются междисциплинарность и интегрированный подход к решению задач проактивного интеллектуального управления логистическими функциями в цепях поставок. Междисциплинарность

проявляется в комплексном использовании накопленного положительного опыта в логистическом менеджменте, в положительном эффекте, получаемом при совместном применении разнородных моделей интегрированной логистики, методов теории управления и средств современных ИИТ. Эти ИИТ позволяют накопленные слабо структурированные знания, накопленные в области логистического менеджмента конструктивно использовать при проактивном управлении ЛФ ЦП. Причем указанная интеграция осуществляется на конструктивном уровне в рамках единой системы проактивного интеллектуального управления ЛФ ЦП, что позволяет получить системный эффект, а также новые знания в предметной области.

Исследования, выполненные по данной тематике, проводились при финансовой поддержке ведущих университетов Российской Федерации: СПб ГПУ (мероприятие 6.1.1), ИТМО (субсидия 074–U01), грантов РФФИ (№№ 16-07-00779, 16-08-00510, 16-08-01277, 16-29-09482-офи-м, 17-08-00797, 17-06-00108, 17-01-00139, 17-20-01214, 17-29-07073- офи-м, 18-07-01272, 18-08-01505), Госзадания Министерства образования и науки РФ №2.3135.2017/4.6, в рамках бюджетных тем №№0073–2018–0003 (№ гос. регистр. АААА-А16-116030250074–1), Международного проекта ERASMUS+, Capacity building in higher education, №73751-EPP-1-2016-1-DE-EPPKA2-CBHE-JP, Новые стратегии обучения инженеров с использованием сред визуального моделирования и открытых учебных платформ

ЛИТЕРАТУРА

1. Лукинский В. С., Лукинский В. В., Плетнева Н. Г., Воробьева Н. И., Маевский А. Г. Управление запасами в цепях поставок. Учебник и практикум для бакалавриата и магистратуры. В 2-х частях. /под общ. ред.: В. С. Лукинский. Ч. 1, 2. М.: Юрайт, 2017.
2. Лукинский В.С., Лукинский В.В., Плетнева Н.Г. Логистика и управление цепями поставок. Учебник и практикум для академического бакалавриата. М.: Юрайт, 2016.
3. Lukinskiy V., Panova Y. Analysis and synthesis of the designed logistics systems, in: Логистика: современные тенденции развития: материалы XVI Междунар. науч.-практ. конф. 6, 7 апреля 2017 г. / Отв. ред.: В. С. Лукинский. СПб. : ГУМРФ имени адмирала С.О. Макарова, 2017. P. 249-252.Мозг фирмы. М.:УРСС, 2005. 315 с.
4. Искандеров Ю.М., Яковлева А.В. Использование информационной системы для обеспечения экспертизы качества инновационных процессов. Серия: Экономика и экологический менеджмент. 2015. №3. С. 210-214. eISSN: 2310-1172.
5. Искандеров Ю.М., Дорошенко В.И. Организация транспортно-технологических процессов на основе интегрированных информационных систем. *Сборник статей Международной научно-практической конференции "Новая экономика и основные направления ее формирования". г.СПб., ФГАОУ ВО "Санкт-Петербургский политехнический университет Петра Великого"*. 2016. С.53-62.
6. Искандеров Ю.М. Построение моделей интегрированной информационной системы транспортной логистики на основе мультиагентных технологий. *Сборник статей Международной научно-практической конференции "Новая экономика и основные направления ее формирования". г.СПб, ФГАОУ ВО "Санкт-Петербургский политехнический университет Петра Великого".* 2016. С.62-69.
7. Искандеров Ю.М. Использование интегрированных информационных систем транспортной логистики при организации взаимодействия железных дорог и морских портов. *Сборник статей международной научно-практической конференции «Рынок и государство в системе экономических отношений». г. СПб., ФГАОУ ВО "Санкт-Петербургский политехнический университет Петра Великого"*. 2016. С. 51-57.
8. Соколов Б.В., Некрасов А.Г., Миротин Л.Б. Разработка и реализация методологии и методик совместного многокритериального синтеза и адаптивного управления созданием, применением и развитием функционально-устойчивых интегрированных транспортно-логистических и информационных систем нового поколения. *Вестник транспорта.* 2011. №6. С.25-30.
9. Охтилев М.Ю., Соколов Б.В., Юсупов Р.М. Интеллектуальные технологии мониторинга и управления структурной динамикой сложных технических объектов. М.: Наука, 2006. 410 с.
10. Охтилев М.Ю., Соколов Б.В., Юсупов Р.М. Интеллектуальные технологии мониторинга и управления структурной динамикой сложных технических объектов. М.: Наука, 2006. 410 с.
11. Некрасов А.Г., Синицына А.С. Логистический инжиниринг как инструмент интеграции логистических систем. *Логистика*. 2016. №12. С. 40-45.
12. Некрасов А.Г, Синицына А.С. Трансформация интегрированных транспортно-логистических систем в цифровую индустрию. *Логистика*. 2017. №8. С. 36-41.
13. Некрасов А.Г., Соколов Б.В., Атаев К.И. Система управления жизненным циклом (трансформация в цифровую инфраструктуру). Учебное пособие. М.: Техполиграфцентр, 2017. 155 с.

V.S. Lykinskiy (Higher School of Economics, St.Petersburg), **Y.M. Iskanderov, B. V. Sokolov.** (St.Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St.Petersburg), **A.G. Nekrasov** (Moscow Automobile and Road Construction State Technical University (MADI), Moscow)

The report proposes an approach aimed at developing and researching the scientific and methodological support of the processes of proactive intelligent management of logistics functions in supply chains based on the introduction and use of relevant intellectual systems that allow integrating (combining) the fundamental and applied results obtained at present in logistical management and the theory of management of complex systems. The characteristic of the current state of the problem being solved is given and the substantiation of the chosen approach is stated. The report presents the results that allow to form scientific and methodical support for the use of intelligent information technologies in logistics systems for the purpose of their effective management, as well as the prospects for the development of research that provides the creation of new technologies for social and economic development in the subject domain.

МОДЕЛИРОВАНИЕ ПРОТОКОЛОВ МАРШРУТИЗАЦИИ MANET СЕТЕЙ

А. Ю. Дорогов[1]
(*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»,*
С.-Петербург, 197376, ул. Профессора Попова 5,
тел. 812-234-37-98, vaksa2006@yandex.ru)

Аннотация

Рассмотрены имитационные модели протоколов маршрутизации FSR, GSRu зондового протокола для использования в беспроводных сетях с динамически изменяемой топологией. Описаны принципы построения имитационной модели в программной среде МАТЛАБ. Представлены результаты моделирования сети для различных вариантов параметров протоколов. Показана необходимость оптимального выбора параметров настройки протоколов для уменьшения времени сходимости алгоритмов. Исследованы особенности работы зондового протокола.

Ключевые слова: маршрутизация, протокол, имитационная модель, беспроводные сети, FSR, MANET.

Введение

Перспективным подклассом беспроводных мобильных телекоммуникационных сетей являются ad hoc сети (MANET – Mobile Ad hoc NETworks) – самоорганизующиеся сети с динамически изменяющейся топологией. Классическими областями применения подобных радиосетей являются боевые действия, чрезвычайные ситуации, мобильные роботы, сенсорная телеметрия, мобильные конференции и т. п. Особенностями мобильных сетей являются: динамичная топология; ненадежность и динамичность радиоресурса, коллективный характер его использования; ограниченность и неоднородность ресурсов узлов (энергоемкость батарей, производительность процессоров, объем памяти и т.д.); ограниченная безопасность и др. [1]. Каждый узел в ad hoc сети работает одновременно как терминальный узел и как маршрутизатор. Поэтому ключевой задачей в MANET сетях является задача маршрутизации.

Маршрутизация принадлежит сетевому уровню относительно семиуровневой ISO/OSI модели. Протоколы маршрутизации используются для обеспечения узлов информацией, необходимой для отправки пакетов следующему узлу на пути от источника до получателя по всем маршрутам.

Основными требованиями к протоколам маршрутизации мобильных сетейявляются: децентрализованное функционирование; минимальная загрузка сети служебной информацией; отсутствие зацикливания маршрутов; быстрая сходимость; построение маршрута заданного качества; поддержка однонаправленных каналов и др. В данной работе представлены методы и результаты трех протоколов маршрутизации. Моделирование проводилось для сети, состоящей из 32 узлов. Топология сети и распределение степеней связанности вершин показаны на рис. 1. При моделировании предполагалось, что все каналы связи имеют задержки, определяемые скоростью передачи данных и длиной передаваемых пакетов. Разработанные модели позволяет оценить динамику протокола, время сходимости алгоритма маршрутизации, служебный трафик, средние и пиковые нагрузки в сети, размеры очередей маршрутизаторов и выбрать оптимальные настроечные параметры алгоритма.

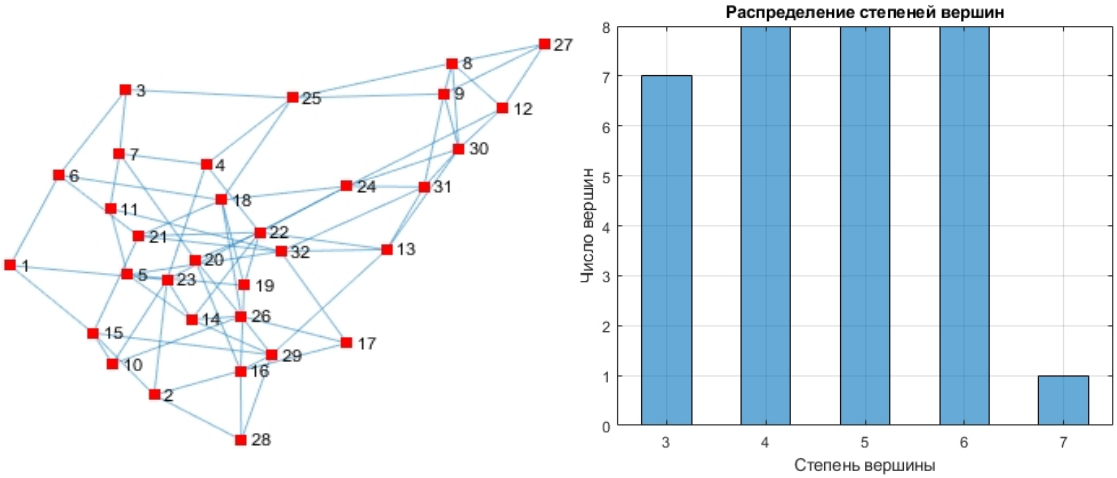
Методы моделирования

Оценку характеристик протоколов осуществляют как правило средствами имитационного моделирования. На рынке представлены достаточно много специализированных платформ имитационного моделирования для телекоммуникационных сетей, к ним относятся: Prophesy, Modeler, NetMarker XA, SES/Stratigizer, Arena, NS-3 и др. Данные средства содержат встроенные библиотеки стандартных протоколов семиуровневой модели ISO/OSI и используются для моделирования различных вариантов типовых сетей. Модели сетей создаются достаточно быстро, как правило, с помощью визуального проектирования и имеют развитые возможности

[1] Доктор технических наук, профессор.

по отображению результатов. Однако для моделирования новых протоколов в рамках данных платформ требуются значительные усилия, связанные с расширением библиотек алгоритмов. В этой ситуации универсальные моделирующие средства такие как МАТЛАБ обладают преимуществом по оперативности разработки частных моделей.

Создание имитационных моделей протоколов маршрутизации в среде МАТЛАБ может быть основано на принципе «календаря событий». Календарь представляет собой общую очередь в которой размещаются служебные пакеты, поступающие от всех узлов сети. Протоколы маршрутизации осуществляют пересылку служебных пакетов между смежными узлами, которые содержат информацию о топологии сети.В моделирующей программе каждый пакет содержит адресную часть, служебную информацию и метку времени, определяющую время наступления события связанного с началом обработки пакета на целевом узле. Метка времени в порождаемом пакетеопределяется скоростью распространения пакета через сегмент сети, а такжемоментом завершения передачи последнего пакета выходной очереди маршрутизатора на том же сегменте. Общая очередь календаря событий обрабатывается во временной шкале по дисциплине «первый пришел-первый обработан».



а) б)

Рис. 1. Топология тестовой сети и распределение степеней связанности вершин

Моделирование FSR протокола

FSR (Fisheye State Routing) -иерархичный проактивный протокол маршрутизации [2], целью создания которого была попытка уменьшить объем рассылаемой по сети служебной информации засчет использования биологической концепции «рыбьего глаза».При реализации данной-концепции каждый узел сети осуществляет периодические широковещательные рассылки ко всем своим ближайшим соседям,причем для сокращения служебного трафика топологическая информацияоб узлах сети,известная текущему узлу, включается в рассылку только частично, в зависимости от удаленности узлов.Чем дальше находится узел от текущего узла, тем реже информация о нем включается в обновляющую рассылку.

Топологическая информация текущего узла представляет собой таблицу, где каждому известному узлу сети соответствует одна строка, содержащая актуальный список его ближайших соседей, время последнего обновления и порядковый номер обновления.Порядковые номера используются при заменах записей. В момент широковещательной рассылки топологической информации, узел сначала проверяет по таймауту записи в своей топологической таблице,и если топологическая информация об узле не обновлялась в течении интервала таймаута, то она уничтожается, это правило позволяет отслеживать изменения топологии сети. Временной интервал таймаута устанавливается трехкратным по отношению к интервалу обновления.Широковещательное обновляющее сообщение к ближайшим узлам представляет собой набор записей, выделенных из топологической таблицы текущего узла.

Когда узел получает обновляющеесообщение, он сначала проверяет узел отправителя на принадлежность к своему списку состояний связей. Если передающий узел является новым ближайшим соседом, он будет добавлен в этот список. В противном случае он будет обновлен по метке времени в списке соседей. Узел затем обрабатывает информацию о связях, содержащуюся в сообщении обновления по следующему правилу:

1. Если это новый узел, ранее неизвестный текущему узлу то будет создана новая запись в топологической таблице и заполнена соответствующей информацией.
2. В противном случае только самое последнее назначение копируется к топологической таблице. Т.е. если какая-нибудь запись во входном сообщении имеет больший порядковый номер, относящийся к узлуjпо сравнению с номером, сохраняемым в топологической таблице текущего узла, локальная запись будет заменена входящей записью.
3. Для входящейзаписи не попадающего в два описанных случая, т.е. если еепорядковыйномер меньше чем предварительный порядковый номер сохраняемой записив текущей топологической таблиц узла, эта запись должна быть послана при следующем обновлении.

Принятый пакет после извлечения информации уничтожается. Во всех случаях, когда топологическая запись добавляется или корректируется,в топологической таблице фиксируется текущее время обновления. После того как все входящие записи проверены, таблица маршрутов вычисляется заново, но только в том случае если обнаружены какие-либо изменения в топологической таблице.Для расчета маршрутов используется алгоритм Дейкстры, мерой расстояния может служить либо минимальное число сегментов сети на маршруте между узлами, либо минимальная временная задержка пересылки пакета по многосегментному маршруту. По степени удаленности от текущего узла все узлы разделяются наFSR-зоны.Топологическая информация об узлах, принадлежащихFSR-зоне *i*включаетсяв пакеты рассылки текущего узла с одинаковой частотой. Критерием сходимости алгоритма маршрутизации в модели могут служить тождественность восстановленных на каждом узле топологий сети к априорно известной топологии.

Характеристики работы алгоритма FSR-маршрутизации определяются настраиваемыми временными параметрами. Методик расчета этих параметровв настоящее время не существует. Рекомендуемые значения далеки от оптимальных, что приводит к необходимости моделирования протокола для конкретной топологии сети.

Результаты моделированияFSRпротокола

Нормативный документ по FSRпротоколу [3] рекомендует использоватьв сети две FSRзоны: ближнею и дальнею с параметрами протокола представленными в колонке «Рекомендация» таблицы 1.

Таблица 1.

Параметры моделирования протокола FSR

Параметр	Обозначение параметра	Рекомендация сек.	Эксперимент сек.
Интервал рассылки для узлов ближней зоны	IntraScope_Interval	5	1
Интервал рассылки сообщений для удаленных узлов	IntreScope_Interval	15	3
Таймаут хранения топологической информации	NEIGHBOR_TIMEOUT	15	3

Документ [3] рекомендует также использовать для передачи топологической информации IPпакеты. Заголовочная часть IP пакета содержит 32 разрядные адреса узлов источника и получателя, полную длину пакета в байтах, контрольную сумму и другие поля, необходимые для поддержки IP-протокола, часть этих полей опциональна. Минимальная длина заголовка 20 байт. Информационная часть пакета определена документом [3] и представлена в таблице 2.

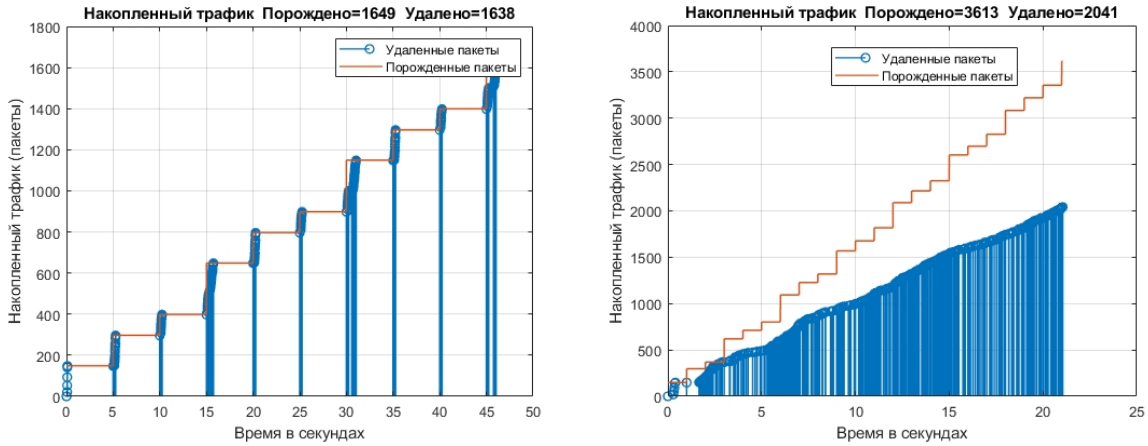
Таблица 2.

Формат информационной части пакета FSR - протокола

Длина пакета(16)	Резерв(16)
Адрес узла1(32)	
Порядковый номер обновления узла1(24)	N1-число соседних узлов(8)
Адрес соседнего узла 1(32)	
-//-	
Адрес соседнего узла N1(32)	
Адрес узла 2(32)	
Порядковый номер обновления узла2(24)	N2 –число соседей узла 2(8)
Адрес соседнего узла1(32)	
-//-	
Адрес соседнего узла N2(32)	
-//-	

Числа в скобках указывают размер полей в битах. Пакет состоит из записей, где каждая запись содержит адреса ближайших соседей для узла сети, порядковый номер обновления и число соседних узлов.Представленный выше формат не содержит адресные заголовки пакета и управляющие поля IPпакета. По топологии сети показанной на рис 1а)не сложнорассчитать максимальный размер пакета топологической информации для моделируемой сети, при 32 разрядных адресах с учетом 20 байтного заголовка IPпакета он будет равен 744 байта.

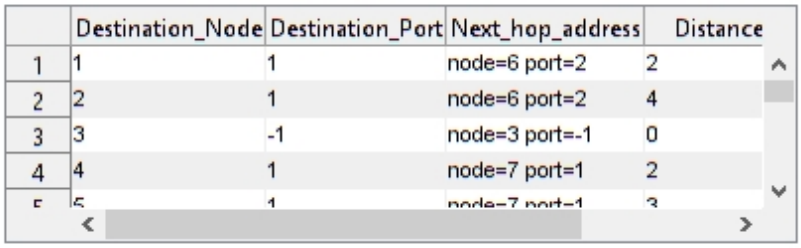
На рис. 2 показана динамика трафика служебных пакетов протокола маршрутизациипри скорости в каналах 1200 бит/сек.Ступенчатые линии соответствуют накопленному трафику порождаемых пакетов. Длина ступени определяется параметром IntraScope_Interval – интервалом рассылки сообщений в ближней зоне. Вертикальные стробы на графике соответствуют накопленному трафику удаляемых пакетов.



а) Рекомендация б) Эксперимент

Рис. 2 Динамика служебного трафика FSR-маршрутизации

Интервал рассылки сообщений для дальней зоны (IntreScope_Interval) в три раза больше чем для ближней зоны, поэтому каждая третья ступень графиков выше двух предыдущих, и соответственно гуще будут стробы удаляемых пакетов (см. рис. 2б)).Получив пакеты обновления, узлы осуществляют восстановление топологии сети и расчет дистанций ко всем узлам сети по алгоритму Дейкстры. На основе данных расчетов строится маршрутно-адресная таблица, определяющая оптимальное продвижение информационных пакетов по сети. Пример таблицы для узла 3 показан на рис. 3.



	Destination_Node	Destination_Port	Next_hop_address	Distance
1	1	1	node=6 port=2	2
2	2	1	node=6 port=2	4
3	3	-1	node=3 port=-1	0
4	4	1	node=7 port=1	2
5	5	1	node=7 port=1	3

Рис. 3. Маршрутно-адресная таблица

Критериями сходимости алгоритма маршрутизации в модели служит тождественность топологий сети построенных на каждом узле с априорно известной топологией сети.Моделирование проводилось для двух вариантов параметров (столбцы «Рекомендация», «Эксперимент» в таблице 1). При расчете длины обновляющего сообщения использовался формат, представленный в таблице 2 с добавкой 20 байтного заголовка IPпакета.В таблице3 приведены результаты моделирования по двум вариантам значений параметров.Значения служебной нагрузки на сетьуказанные в таблице вычисляются на интервале сходимости алгоритма маршрутизации и нормируются к пропускной способности сети. Расчетная пропускная способность данной сети равна 177600 бит/сек.

Т а б л и ц а 3.

Результаты моделирования протокола FSR

Вариант модели	Время сходимости сек.	Служебный трафик бит.	Средний размер пакета байт	Нагрузка на сеть	Пиковая нагрузка	Max. Размер очереди
Рекомендуемые параметры						
FSR IP 1200 бит/сек	50.8	3193344	239	0.354	0.406	2
FSR IP 2400 бит/сек	47.9	3087904	237	0.182	0.139	2
FSR IP 4800 бит/сек	46.4	3093696	237	0.093	0.099	2
FSR IP 9600 бит/сек	45.7	3093696	237	0.047	0.050	2
FSR IP 19200 бит/сек	45.4	3093696	237	0.024	0.025	2
Экспериментальные параметры						
FSR IP 1200 бит/сек	21.1	3259616	200	0.872	1.87	14
FSR IP 2400 бит/сек	11.9	2664320	220	0.631	0.989	4
FSR IP 4800 бит/сек	10.4	3134144	236	0.422	0.501	2
FSR IP 9600 бит/сек	9.72	3087904	237	0.224	0.249	2
FSR IP 19200 бит/сек	9.36	3093696	237	0.116	0.124	2

Рис 2. Демонстрирует два крайних случая выбора настроечных параметров протокола маршрутизации. На рис 2а) время обработки служебных пакетов занимает малую долю интервала рассылки. Поэтому без ущерба работы протокола настроечные параметры могут быть уменьшены, что позволит уменьшить время сходимости алгоритма. Время обработки пакетов включает в себя и время доставки на соседний узел, с увеличением скорости передачи данных время доставки уменьшается поэтому степень уменьшения значений параметров может быть больше. На рис 2б) показан противоположный случай, когда параметры выбраны слишком малыми и при низкой канальной скорости (1200 бит/сек) на интервале рассылки не хватает времени для обработки пакетов, в результате образуются очереди на обработку и время сходимости алгоритма увеличивается.

Результаты моделированияGSRпротокола

Протокол GSR (GlobalStateRouting) является вариантом FSR протокола с единственной FSR-зоной. В этом случае не существует различий между узлами дальней и ближней зоны. Обновление топологической информации со смежными узлами в этом случае определяется константами представленными в таблице 4.

Таблица 4

Параметры моделирования протокола FSR

Параметр	Обозначение параметра	Рекомендация сек.	Эксперимент сек.
Интервал рассылки для узлов ближней зоны	IntraScope_Interval	5	1
Таймаут хранения топологической информации	NEIGHBOR_TIMEOUT	15	3

Моделирование выполнялось для двух наборов параметров «Рекомендация» и «Эксперимент». Результаты моделирования представлены в таблице 5.

Таблица 5.

Результаты моделирования протокола GSR

Вариант модели	Время сходимости сек.	Служебный трафик бит.	Средний размер пакета байт	Нагрузка на сеть	Пиковая нагрузка	Max. Размер очереди
Рекомендуемые параметры						
GSR IP 1200 бит/сек	30.8	2611744	422	0.477	0.783	2
GSR IP 2400 бит/сек	22.9	2153472	422	0.285	0.402	1
GSR IP 4800 бит/сек	21.5	2153472	422	0.141	0.201	1
GSR IP 9600 бит/сек	20.7	2153472	422	0.073	0.101	1
GSR IP 19200 бит/сек	20.4	2153472	422	0.037	0.05	1
Экспериментальные параметры						
GSR IP 1200 бит/сек	51.0	8384328	372	0.926	3.00	36
GSR IP 2400 бит/сек	11.25	2965152	349	0.742	1.55	6
GSR IP 4800 бит/сек	6.5	2704448	441	0.586	1.06	3
GSR IP 9600 бит/сек	4.7	2153472	422	0.321	0.503	1
FSR IP 19200 бит/сек	4.4	2153472	422	0.174	0.251	1

Моделирование зондового протокола

Суть метода зондовой маршрутизации [4] заключается в следующем: с любого узла начинается рассылка к смежным узлам коротких пакетов, называемыми зондами, которые накапливают в себе информацию о состоянии сети. Каждый зонд содержит битовую карту, длина которой равна максимально-возможному числу узлов сети. При прохождении узла в битовой карте зонда устанавливается единица в позиции, определяемой номером узла. При получении зонда узел анализирует информацию, содержащуюся в битовой карте, и по результатам анализа добавляет записи в собственную маршрутно-адресную таблицу. После чего узел проверяет ряд условий для принятия решения о репликации или уничтожении зонда. В случае репликации в зонд добавляется информация о прохождении текущего узла и зонд рассылается по всем исходящим ветвям текущего узла.

Маршрутно-адресные таблицы (матрицы маршрутизации) строятся на каждом узле сети. Число столбцов в матрице равно числу целевых узлов в сети (номер текущего узла из списка целевых узлов исключается, но столбец текущего узла формально присутствует в матрице). Каждый элемент матрицы маршрутизации содержит номер порта текущего узла, указывающего направление через которое следует передавать пакет для целевого узла. При зондовом протоколе матрица маршрутизации содержит столько строк сколько портов имеет текущий узел. Строки упорядочены по приоритетам. В первой строке содержатся направления высшего приоритета, обеспечивающего наиболее быструю доставку пакета к целевым узлам. Во второй строке содержаться направления, обеспечивающие более медленную доставку пакета по сравнению с первой и т.д., таким образом приоритет направлений снижается при увеличении номера строки матриц маршрутизации. Пример построенной матрицы маршрутизации для узла 5 показан на рис. 4.

	Node-1	Node-2	Node-3	Node-4	Node-5
1	2	3	1	1	-1
2	1	2	4	2	-1
3	4	4	3	3	-1
4	3	1	2	4	-1

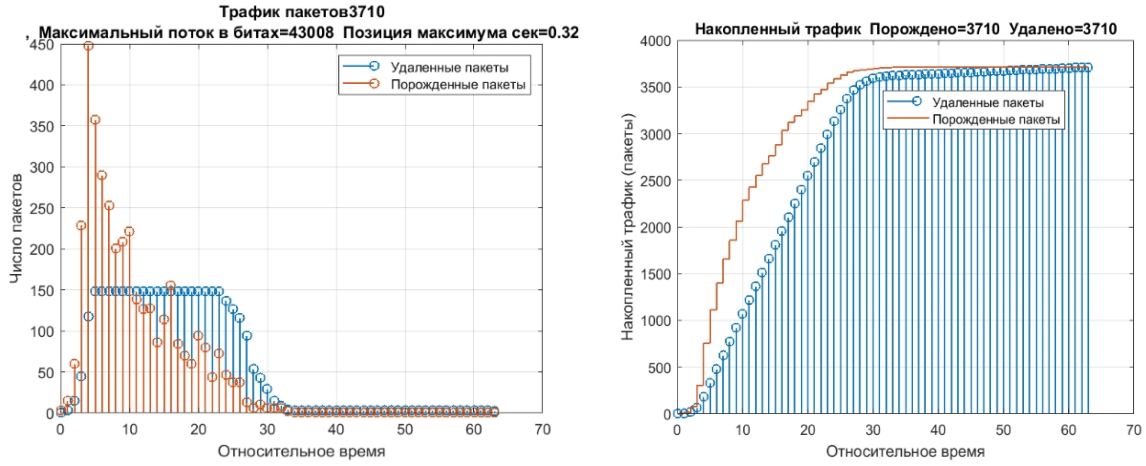
Рис. 4. Матрица маршрутизации

В начале цикла маршрутизации маршрутно-адресная таблица узла заполнена нулями за исключением столбца, соответствующего текущему узлу, который заполнен значением (-1). Если зонд прибыл в узел первым, то заполняется первая строка матрицы маршрутизации, следуя правилу: в элементы строки с номерами узлов, через которые прошел зонд, заносится номер порта, через который зонд был получен на текущем узле. Поступление последующих зондов на текущий узел приводит к заполнению нулевых элементов первой строки, которые не были заполнены ранее, а в случае их занятости последовательно заполняются нулевые элементы второй и последующих строк. Если в момент поступления зонда все элементы матрицы маршрутизации узла заполнены, то зонд уничтожается. Зонд будет также уничтожен, если информация на узле уже содержит информацию, доставленную зондом. В остальных случаях зонд добавляет в свою битовую карту текущий узел и реплицируется по всем исходящим направлениям. В случае обнаружения цикла, т.е. когда информация, доставленная зондом уже содержит номер текущего узла зонд рассылается по всем направлениям кроме того по которому он был получен (для стартового узла это ограничение не действует). Критерием сходимости алгоритма маршрутизации в модели могут служить исчерпание очереди календаря событий.

Результаты моделирования зондового протокола

Моделирование проводилось для сети, показанной на рис. 1а). При моделировании предполагалось, что все каналы связи имеют одинаковые задержки. На рис. 5 показана динамика трафика служебных пакетов при скорости в каналах 1200 бит/сек. Размер зонда выбран равным 12 байт. На приведенных графиках единица относительного времени соответствует длительности пересылки зонда по каналу связи. В качестве стартового узла использовался первый узел. В результате моделирования получено:

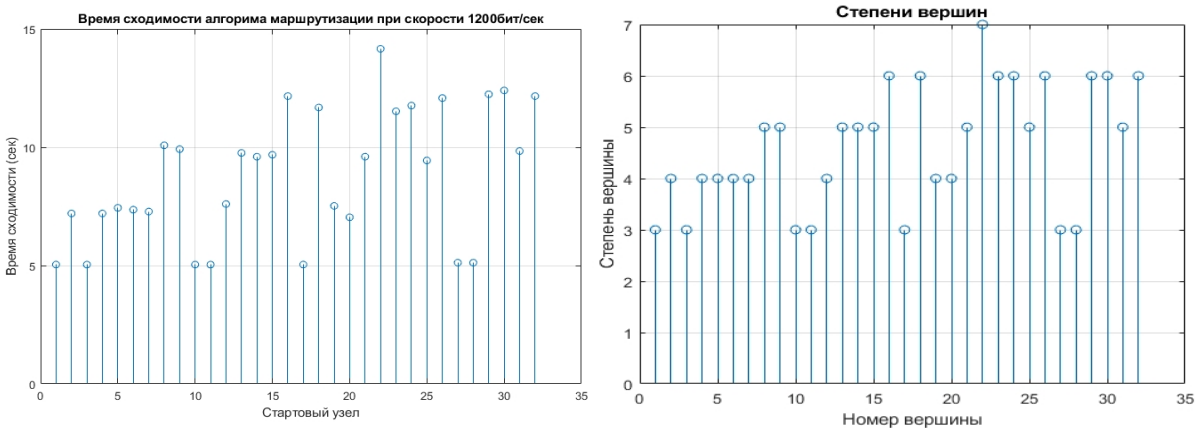
- Полный служебный трафик зондов составил 3710 пакетов, при размере каждого зондового пакета 12 байт.
- Время сходимости алгоритма маршрутизации 5.04 сек при скорости передачи данных 1200 бит/сек.
- Средняя нагрузка на сеть на интервале сходимости 0.398 от пропускной способности сети.
- Пиковая нагрузка на момент максимума трафика равна 1.27 от пропускной способности сети.
- Поскольку пиковая нагрузка превышает пропускную способность возникают очереди на выходе маршрутизаторов. Максимальная длина выходной очереди при данных условиях эксперимента составила 29 пакетов.



а) Динамика трафика б) Накопленный трафик

Рис. 5. Динамика служебного трафика зондового протокола маршрутизации

Для зондового протокола время сходимости алгоритма маршрутизации, и нагрузка сети зависят от стартового узла распространения зондов. На рис. 6а) показана скорость сходимости алгоритма маршрутизации в зависимости от выбора стартового узла.

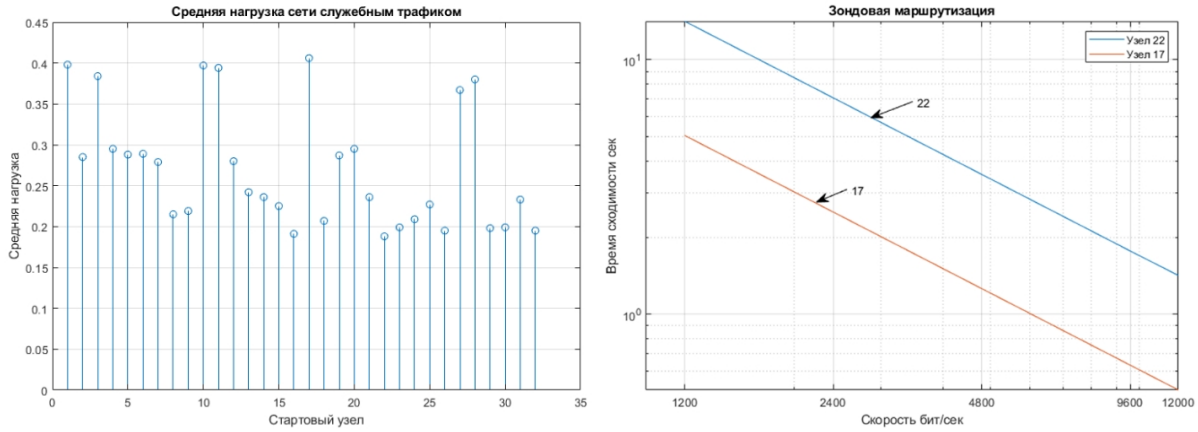


а) б)

Рис. 6. Эффект влияния стартового узла на время сходимости алгоритма

Максимальное время равно 14.16 сек. (узел 22), минимальное время равно 5.04 сек. (узел 17). На рис. 6б) показаны степени вершин исследуемой сети. Сравнение двух графиков показывает, что существует явная зависимость времени сходимости алгоритма от степени стартового узла, чем выше степень стартового узла, тем больше время сходимости алгоритма маршрутизации.

На рис. 7а) представлена зависимость средней нагрузки создаваемой служебным трафиком в зависимости от стартового узла. По оси ординат отложена нагрузка, нормированная к пропускной способности сети.Объем трафикадля стартового узла 17 составил 362976 бит (3781 пакет), для стартового узла 22: 472224 бит (4919 пакетов).



а) б)

Рис. 7. Средняя нагрузка на сеть и время сходимости алгоритма зондовой маршрутизации

Максимальная степень загрузки сети служебным трафиком 0.406 (узел 17) а минимальная 0.188 (узел 22). Из приведенных графиков видно, что чем выше степень загрузки сети служебным трафиком, тем меньше время сходимости алгоритма маршрутизации. Учитывая зависимость времени сходимости алгоритма от степени вершины можно заключить что, степень загрузки служебным трафиком тем выше, чем меньше степень стартовой вершины.

С увеличением скорости передачи данных время сходимости алгоритма маршрутизации линейно уменьшается.Характер зависимости для стартовых узлов 17 и 22 показан на рис 7б). На скорости 9600 бит/сек время сходимости для стартового узла 22 составляет 1.77 сек, а для узла 17 составляет 0.63 сек. Пропускная способность сети линейно увеличивается с увеличением скорости, при этом нормированная нагрузка сети по служебному трафику не изменяется.

Дефекты матриц маршрутизации

В идеальном варианте работы алгоритма зондовой маршрутизации предполагается, что в процессе распространения зондов по сети все элементы матриц маршрутизации будут полностью заполнены. Так и происходит для сетей с малым количеством узлов. Однако результаты моделирования для тестовой сети с 32 узлами показывают, что по завершению алгоритма маршрутизации полного заполнения матриц не происходит – в матрицах маршрутизации остаются не заполненные элементы (дефекты). Число дефектов в матрицах маршрутизации зависит от стартового узла. На рис. 8 показаны количественные оценки суммарных дефектов матриц маршрутизации в зависимости от стартового узла. При расчете оценок, суммирование дефектов выполнялось по всем узлам и нормировалось в процентах к общему объему заполняемых элементов. Из графика видно, что объем дефектов в худшем случае не превосходит 0.5%. Сравнение графика оценки со степенями вершин показывает, что минимальный процент дефектов достигается для стартового узла с максимальным числом портов (максимальной степенью вершины). Дефекты в матрицах маршрутизации концентрируется в строках с наименьшим приоритетом направлений. При практической реализации алгоритма маршрутизации не заполненные позиции матриц, могут быть без ущерба заполнены значениями строки с более высоким приоритетом.



Рис. 8. Дефекты в матрицах маршрутизации

Заключение

Выбор настроечных параметров протокола FSRзависит от топологии и скорости передачи данных в линиях связи. Параметры могут быть определены в результате моделирования. Критерием оптимальной настройки алгоритма по времени сходимости может служить наполненность интервалов рассылки процессами обработки пакетов. Эксперименты показывают, что для сети с данной топологией рекомендуемые параметры могут быть уменьшены в 3-4 раза без нарушения работы протокола. В результате, время сходимости алгоритма маршрутизации можно уменьшить более чем в 4 раза. Для окончательного выбора параметров протокола FSR для сети MANET необходимо выполнить моделирование для всех потенциально возможных вариантов «максимальной топологии» сети.

Протокол зондовой маршрутизации является распределенным, работает в асинхронном режиме и обладает исключительно простой алгоритмической реализацией. Благодаря этим особенностям обеспечивается высокая скорость сходимости алгоритма. Причем время сходимости, в отличии от алгоритма FSR, уменьшается пропорционально увеличению скорости передачи данных. Метод зондовой маршрутизации позволяет построить несколько вариантов маршрутов и упорядочить их по предпочтительности. Протокол маршрутизации учитывает фактические задержки в каналах связи и поэтому может адаптироваться к степени загрузки каналов сети. Протокол обладает очевидной зависимостью времени сходимости алгоритма и степени служебной нагрузки на сеть от выбора стартового узла, что позволяет найти оптимальное решение для конкретной топологии сети.

ЛИТЕРАТУРА

1. Миночкин А. И., Романюк В. А. Маршрутизация в мобильных радиосетях – проблема и пути ее решения // Зв'язок, 2006, № 3.
2. Guangyu Pei, Mario Gerla, Tsu-Wei Chen Fisheye State Routing in Mobile Ad Hoc Networks Workshop on Wireless Networks and Mobile Computing, Taipei, Taiwan, April. 2000.
3. Fisheye State Routing Protocol (FSR) for Ad Hoc Networks. Режим доступа: https://tools.ietf.org/html/draft-ietf-manet-fsr-03.
4. Аванесов М.Ю. Метод зондового управления потоками данных в мультисервисных сетях с быстро меняющейся структурой // Информация и космос, №3, 2007.- с.35-40.

A.Yu.Dorogov (Saint Petersburg State Electrotechnical University “LETI”)

Modeling of Route Protocols for MANET networks

Imitation models of route protocols so as FSR, GSR and probing route protocol for useful in wireless networks with dynamic exchanged topology are considered. The principles of imitation model designing in program environment MATLAB are described. The results of network modeling for different setting variants are represent. Necessity of optimal chose for setting of protocol parameters to decrease convergence algorithm time is shown.Special features of probing protocol is investigated.

ПРИНЦИПЫ И ПОДХОДЫ К ИСПОЛЬЗОВАНИЮ СЕРВИСНО-ОРИЕНТИРОВАННОЙ АРХИТЕКТУРЫ ДЛЯ ФОРМИРОВАНИЯ ИНТЕЛЛЕКТУАЛЬНОГО ИНФОРМАЦИОННОГО ПРОСТРАНСТВА ПРЕДПРИЯТИЯ

В. В. Алферов[1], **В. В. Кондратьев**[2], **Б. В. Соколов**[3], **М. Ю. Охтилев**[4], **В. И. Салухов**[5]
(*Федеральное государственное бюджетное учреждение науки
«Санкт-Петербургский институт информатики и автоматизации» Российской академии наук,
199178, Санкт-Петербург, 14 линия, д. 39,
komplekt@komplekt.gazprom.ru, 3068129@gmail.com, sokolov_boris@inbox.ru,
oxt@mail.ru, vsaluhov@bk.ru*)

Аннотация

В докладе предложена концептуальная и формальная модели единого интеллектуального информационного пространства (ЕИИП) предприятия, основанные на новом подходе к распределению сенсорных, сетевых, вычислительных, информационно-управляющих и сервисных задач между встроенными в одежду субъекта устройствами, мобильными клиентскими устройствами, стационарным сервисным оборудованием, облачными вычислительными и информационными ресурсами. Формализована задача структурно-функционально-параметрического синтеза распределенной киберфизической системы и соответствующего ЕИИП, показана возможность ее решения как методами инженерии знаний, теории онтологий, теории управления сложными логико-динамическими объектами с перестраиваемой структурой, так и методами, разрабатываемыми в исследовании операций и теории распознавания образов. В докладе рассматривается в качестве примера, где необходима реализации предлагаемого подхода к формированию ЕИИП, задача интеграции данных и информации о состоянии объектов на радиационно опасных предприятиях.

Ключевые слова: методы, модели и алгоритмы реконфигурации, бортовая аппаратура.

Введение

На современном предприятии используются различные виды датчиков для получения данных и информации о состоянии различных классов объектов, находящихся на его территории. Указанные датчики входят в состав таких систем, как системы контроля и управления доступом (СКУД), системы видеонаблюдения, системы физического разделения функциональных зон по принципу авторизованного доступа, маркирование работника (бейдж, окраска рабочей одежды, защитной каски), системы оповещения и указания рабочего маршрута и пути безопасной эвакуации и т.д. Кроме этого, наряду с техническими объектами (ТехО),на территории предприятия ведется контроль физического состоянияего работников (как правило, при входе на предприятие), а именно, медицинский и биометрический контроль. Также на территории предприятия ведутся и другие виды контроля, например, радиационный / химический / биологический, контроль обхода маршрута патрулирующими группами (работниками). В большинстве случаев это связано с необходимостью разграничения общей площади цехов, зданий физическими способами – решетками, дверями, шлагбаумами, снабженными датчиками и исполнительными устройствами СКУД. Современные системы распределенного контроля при их комплексировании (RFID, LoraWan/ZigBee Video/CCTV)могут позволить решить дополнительные задачи такого контроля и управления персоналом [1-7]. Так например, «маркирование» рабочей одежды и средств индивидуальной защиты (СИЗ) системами удаленного контроля и наблюдения позволит проводить непрерывное решение следующих задач [4-13]:

1. Контроль биомедицинского состояния работника (температура тела, пульс, температура тела, состояние кожного покрова (например, влажность). Дополнительно возможно контролировать случаи травмирования работника, например, при падении (в том числе с высоты) самого работника или падения на него предметов (датчики удара/акселерометры).

[1] Соискатель.
[2] Соискатель.
[3] Доктор технических наук, профессор, заведующий лабораторией.
[4] Доктор технических наук, профессор, ведущий научный сотрудник.
[5]Кандидат технических наук, доцент, начальник отдела аспирантуры информационно-образовательных технологий и услуг.

2. Контроль передвижения работника по территории промплощадки, административной зоны, зоны приема пищи и отдыха, в том числе и с целью учета его рабочего времени.

3. Непрерывный контроль состояния производственной среды датчиками, находящимися на такой одежде.

4. Возможность поиска и оповещения каждого работника, вызов как работника в предварительно условленное место, так и вызов рабтником такой помощи и/или его доклад о непредвиденных обстоятельствах, возникших при выполнении им его обязанностей.

5. Аварийный останов исполнительных устройств технологического оборудования, стопорение подъемно-транспортных машин и обесточивание шин питания и электрораспределительных устройств технологического оборудования в случаях, когда работник случайно (неавторизованно) попадает в зону возможного травмирования движущими механизмами, поражения электрическим током и другими опасными производственными факторами. Подача предупреждающего сигнала при попадании работника в зону возможного поражения опасных производственных факторов.

Из анализа изложенного материала следует, что в настоящее время на современном предприятии существует огромное количество датчиков и приборов, регистрирующих первичные данные о состоянии как ТехО, так и работников предприятия. Однако для оценивания в целом состояния самого предприятия и проводимых на нем работ, а также о складывающей внешней ситуации, необходимо в оперативном режиме обрабатывать и интегрировать получаемые огромные информационные потоки, чтобы сформировать целостное представление о текущей обстановке.

Для этого, используя современные информационные технологии и системы, сотрудники ИТ служб соответствующего предприятия поэтапно формируют его единое информационное, а в последующем единое интеллектуальное информационное пространство (ЕИП, ЕИИП).

Далее под единым информационным пространством (ЕИП)будем понимать всю совокупность актуальных и полных информационных ресурсов (ИР) с унифицированными правилами создания и потребления, едиными стандартами представления и возможностью непосредственного (прозрачного) компетентного доступа пользователей к этим ресурсам. Основными компонентами ИП являются информационные ресурсы, средства информационного взаимодействия, информационная инфраструктура [4,6,11]. В свою очередь, единое интеллектуальное информационное пространство – это уже физическая среда со встроенными устройствами, обеспечивающая проактивное ненавязчивое контекстное персонифицированное обслуживание и информационную поддержку находящихся в ней пользователей. Интеллектуальное пространство реализуется на основе технологий повсеместных вычислений, повсеместных коммуникаций и интеллектуальных интермодальных интерфейсов [14]. В настоящее время данные технологии одновременное реализуются в киберфизических системах (КФС), которые объединяют два начала – кибернетическое (вычислительно-коммуникационно-управленческое) и физическое, которое определяется теми физическими процессами, с которыми взаимодействуют указанные системы. КФС отличаются наличием интеллекта и способностью взаимодействовать со своим окружением; планировать и адаптировать свое собственное поведение согласно окружающим условиями их ожидаемым изменениям; учиться новым моделям и линиям поведения и быть самоуправляемыми, в том числе реализующими функции контроля, прогнозирования, диагностирования и метрологического самоконтроля. Их повсеместное внедрение приведет к гораздо большим изменениям, чем появление компьютеров и Интернета. В перспективных КФС, наряду с функциями измерения, диагностики и управления, также будут реализованы функции автоматического составления отчетов о состоянии соответствующей подсистемы контролируемого оборудования, в том числе данные о всех возникающих неисправностях; об остатке ресурса изнашиваемых деталей; о ресурсе расходных материалов; загрузке оборудования, режиме его эксплуатации и т.д. Указанные возможности открывают широчайшие перспективы по автоматизации и интеллектуализации как самого цифрового производства продукции, так и ее обслуживания и эксплуатации во время послепродажного функционирования. Важнейшими элементами предлагаемого электронного обслуживания (e-maintenance) будут являться базирующиеся на Web-технологиях мониторинг, тестирование, диагностика, прогнозирование состояния эксплуатируемых изделий, реконфигурация их структур в случае возникновения аварийных или нештатных

ситуаций и отсутствия необходимых резервов, администрирование, осуществляемые дистанционно.

В качестве целей создания ЕИПобычно полагают [9-14]: расширение возможностей пользователей по прозрачному, компетентному доступу к пространственно-распределенным информационным ресурсам (ИР); расширение возможностей пользователей по распространению, хранению, визуализации и обработки ИР; обеспечение полноты, актуальности и точности ИР, хранимых и потребляемых пользователями; предоставление пользователям новых возможностей по производству знаний. Обобщенная структурно-функциональная схема единого информационного пространства, которая может быть положена в основу построения ЕИП представлена на рисунке 1. Основу ЕИП составляют метаданные (мета-схема ЕИП), представляющие собой формальное описание обобщенной информационной модели данных и знаний, циркулирующих в ЕИП. Метаданные используются для описания и автоматизированного анализа содержимого ресурса, построения поисковых индексов и позволяют обеспечить высокую точность и эффективность выбора разнотипной информации.

Мета-схема ЕИП определяет термины и семантику описания и применения обобщённой информационной модели и должна обеспечивать [14]: описание объектов инфраструктуры ЕИП с перечнем атрибутов и отношений между объектами; ***синтаксическую интероперабельность*** - описание типов данных, форматов и моделей различных информационных ресурсов на основе согласованных унифицированных подходов и стандартов; ***семантическую интероперабельность*** - создание и согласование стандартных прикладных профилей метаданных и онтологий, относящихся к различным предметным областям (ПрО), что позволяет упростить интеграцию разнообразных систем, автоматизацию обмена метаданными, их обработку и преобразование; ***открытость*** - доступ к соответствующей информации по описаниям (мета-элементам); ***расширяемость*** - возможность детализации описаний и добавление новых метаданных; ***уникальную идентификацию*** информации и объектов инфраструктуры - возможность установления взаимосвязей между ресурсами разных информационных источников распределенной среды, способность объединять связанные данные отдельных репозиториев в виртуально-единые ресурсы; ***возможности интеграции информации*** (при расширении/добавлении объектов ЕИП) на основе использования существующих информационных стандартов.

Более того, инфраструктурные компоненты ЕИП должны обеспечивать решение следующих задач [14]: организация обмена информацией со смежными системами в согласованных форматах по стандартным или специализированным протоколам; эффективное размещение и поиск информации в распределенной среде; создание и изменение описаний информационных объектов с возможностью расширения репозитария описаний по мере увеличения данных и знаний об этих объектах с адекватным отображением их в структуре баз данных (БД); сопровождение ЕИП (проверка целостности, коррекция ошибок, задачи репликации, создание резервных копий всего ЕИП или его частей, выгрузка, загрузка и сжатие БД, их защита и т.п.); выбор и обработка данных по группам взаимосвязанных объектов; миграция информации при изменении (расширении) инфраструктуры ЕИП; выбор и обработка связанных данных по нескольким информационным объектам.

Опыт показывает, что создание открытого ЕИП начинается с выделения узла, координирующего работу сети (узел-сервер). Переменная топология распределенных БД (РБД) поддерживается за счет трехуровневой самоорганизации сети: на первом уровне фиксируется появление (исчезновение) соседних узлов (клиентов) и оценивается время активности (неактивности); на втором уровне оцениваются связи между узлами; на третьем уровне по оцененным характеристикам осуществляется кластеризация сети.

Для организации распределенных вычислений в составе ЕИП должны быть такие компоненты, как сетевое программное обеспечение (ПрО) и ПрО компонентных вычислений. Данные компоненты позволяют сформировать сетевую среду, каждый элемент которой решает часть задач (из общего перечня задач, возложенных на ПрО), обусловленных его целевым применением и территориальным расположением. При этом весь перечень задач динамически распределяется по соответствующим элементам сетевой среды. Управление вычислительным процессом в рамках такой распределенной сети производится с использованием формальной парал-

лельной модели вычислений на основе пространства измеряемых и вычисляемых параметров, представляющего собой виртуальное распределенное (компонентное) пространство.

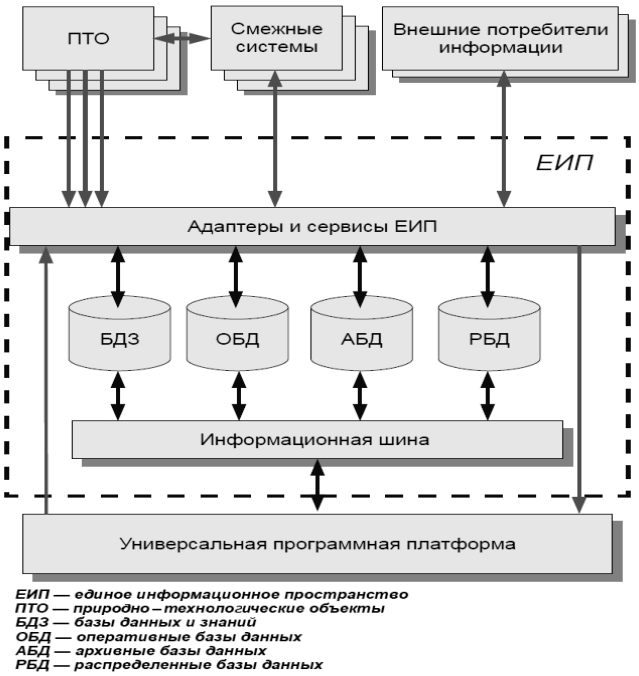


Рис. 1. Обобщенная структурно-функциональная схема ЕИП

В дальнейшем для определённости будем предполагать, что перед КФС, являющейся составной частью ЕИП (ЕИИП), стоит двойственная задача [14-16]:

1. Формирование в заданных областях физического пространства таких условий («информационных полей»), при которых каждый элемент (узел) КФС, находящийся в указанных областях, может определять своё местоположение, обмениваться информацией друг с другом, определять и оценивать состояние внешней среды (пользователя);
2. Создание и поддержание в пространстве такой структуры КФС, которая позволит за заданное (либо минимальное) время осуществлять непосредственно вещественное, либо энергетическое взаимодействие с пользователем, в ходе которого будет выполняться целевая задача, стоящая перед системой.

Следует подчеркнуть, что задача создания вышеперечисленных условий может ставиться как для всего пространства (глобально), так и для какой-либо его части (локально), на всём заданном интервале времени (непрерывно), либо в какие-то дискретные моменты времени, интервалы времени (дискретно). При этом может дополнительно задаваться набор количественных параметров, характеризующих те или иные условия (информационные поля).

В качестве элементов КФС рассматриваются мобильные/носимые, стационарные и передвижными комплекты аппаратно-программных средств сбора, передачи, обработки и хранения данных и информации, а также их облачные компоненты. Основным подсистемам КФС могут быть поставлены в соответствие четыре процесса (вида функционирования): взаимодействие, функционирование целевых и обеспечивающих аппаратно-программных средств, движение, расход и (или) пополнение ресурсов. Рассмотрим формальную постановку задачи формирования ЕИИП с использованием КФС.

Формальная постановка и анализ возможных путей решения задачи формирования единого информационного интеллектуального пространства на предприятии

Целесообразно рассмотреть постановку и решение всех перечисленных задач с единых позиций. Для этого в докладе предлагается концептуальная и формальная модели единого интел-

лектуального информационного пространства (ЕИИП), основанная на новом подходе к распределению сенсорных, сетевых, вычислительных, информационно-управляющих и сервисных задач между встроенными в одежду субъекта устройствами, мобильными клиентскими устройствами, стационарным сервисным оборудованием, облачными вычислительными и информационными ресурсами.

В настоящее время существует большое разнообразие подходов к решению задач формирования как ЕИП, так ЕИИП. Основной недостаток данных подходов состоит в их разобщенном описании и решении соответствующих системотехнических задач. При этом на практике очень трудно согласовать указанные решения, полученные с использованием различных классов моделей, методов и алгоритмов. В рамках разрабатываемой авторами данного доклада теории проактивного управления сложными объектами (СлО), к числу которых относится и современное промышленное предприятие, удалось все перечисленные выше задачи формально описать и исследовать с единых методологических и методических позиций в рамках более общей проблемы структурно-функционального синтеза. В этом случае проблема многокритериального выбора эффективных стратегий и программ формирования как ЕИП, так и ЕИИП может быть формально представлена в следующем виде: необходимо разработать принципы, подходы, модели, методы, алгоритмы, позволяющие находить такие $<U_*^t,\ S_\delta^{*t_f}>$, при которых выполняются следующие условия

$$
\begin{gathered}
J_\theta\left(X_\chi^t,\Gamma_\chi^t,Z_\chi^t,F_{<\chi,\chi'>}^t,\Pi_{(\tilde\delta,\tilde{\tilde\delta})}^t,t\in(t_0,t_f]\right)\to \underset{<U^t,S_\delta^{t_f}>\in\Delta_g}{extr},\\
\Delta_g=\left\{<U^t,S_\delta^{t_f}>\ \middle|\ R_\theta\left(X_\chi^t,\Gamma_\chi^t,Z_\chi^t,F_{<\chi,\chi'>}^t,\Pi_{(\tilde\delta,\tilde{\tilde\delta})}^t\right)\le\tilde R_g;\right.\\
\left.U^t=\Pi_{<\delta_1,\delta_2>}^{t1}\circ\Pi_{<\delta_2,\delta_3>}^{t2}\Pi_{<\tilde\delta,\tilde{\tilde\delta}>}^{t2};\ \chi\in\mathbf{B}\right\}
\end{gathered}
\qquad (1)
$$

где δ – индекс, характеризующий различные типы структур, используемых для формирования как ЕИП (ЕИИП) и входящих в состав соответствующих КФС, $\in\{$Топ, Фун, Тех, ПМО, ИО, Ор$\}$ – множество индексов, соответствующих топологической, функциональной, технической структурам, структурам программно-математического и информационного обеспечения (ПМО, ИО), организационной структуре КФС, t∈T– множество моментов времени; $X_\chi^t=\{x_{\chi l}^t, l\in L_\chi\}$ – множество элементов, входящих в состав структуры динамического альтернативного системного графа (ДАСГ) G_χ^t (множество вершин ДАСГ), помощью которого задается управляемая структурная динамика соответствующих КФС в момент времени t; $\Gamma_{\chi'}^t=\{\gamma_{<\chi,l,l'>}^t, l,l'\in L_\chi\}$ – множество дуг ДАСГ типа G_χ^t, отражающих взаимосвязи между его элементами в момент времени t; $Z_\chi^t=\{z_{<\chi,l,l'>}^t, l,l'\in L_\chi\}$ – множество значений параметров, количественно характеризующих взаимосвязь соответствующих элементов ДАСГ. $F_{<\chi,\chi'>}^t$ - отображения различных структур ЦП и ИС друг на друга в момент времени t; $\Pi_{<\tilde\delta,\tilde{\tilde\delta}>}^t$ - операция δкомпозиции многоструктурных макросостояний КФС с номерами $\tilde\delta,\tilde{\tilde\delta}$ в момент времени t; Ut – управляющие (программные и в реальном масштабе времени) воздействия, позволяющие синтезировать как структуры КФС, так и самих ЕИП (ЕИИП); Jθ – стоимостные, временные, ресурсные показатели, характеризующие качество ЕИП (ЕИИП), q ∈ Q = {1,…,l} – множество номеров показателей; Δ_g – множество динамических альтернатив (множество структур и параметров КФС, множество программ их функционирования); B – множество номеров пространственно-временных, технических и технологических ограничений, определяющих процессы функционирования КФС для различных сценариев возмущающих воздействий; $\tilde R_g$ – заданные величины; T = (t0, tf] – интервал времени, на котором синтезируются ЕИП (ЕИИП)[14]. Особенностью модели вида (1) является то, что с ее помощью можно конструктивно описать и решить традиционные задачи управления ЖЦ КФС(в рамках технологий PLM), но и задачи проактивного управления ЖЦ КФС уже в рамках принципиально новой концепции управления ЖЦ указанных объектов, базирующейся на технологии CL2M.

На первом шаге решения предложенной обобщённой проблемы многокритериального выбора эффективных стратегий и программ формирования как ЕИП, так и ЕИИП необходимо прежде всего рассмотреть вопрос формирования методологических основ её исследования.

Масштабность и сложность рассматриваемой в этой связи проблемы требует формирования соответствующей методологии, в качестве которой должна быть выбрана методология современного обобщённого системного анализа [14-17], представляющего из себя одно из главных направлений реализации системного подхода, в рамках которого на основе гармоничного сочетания формально-математических и логико-эвристических методов осуществляется конструктивное решение разнородных и разноуровневых задач анализа и синтеза как ЕИП, так ЕИИП на различных этапах их жизненного цикла.

Конструктивное решение рассматриваемой проблемы поиска и выбора наилучших вариантов создания и развития как ЕИП, так ЕИИП предполагает, во-первых, построение соответствующего полимодельного комплекса, описывающего различные аспекты их жизненного цикла, во-вторых, разработку методов, алгоритмов и методик многокритериального синтеза структуры перспективной информационной системы (ИС), в рамках которой происходит формирование ЕИП (ЕИИП), и, в-третьих, разработку многоэтапной интерактивной процедуры поиска решения задачи синтеза ИС и программ ее развития.

Анализ многочисленных публикаций по проблемам комплексного (системного) моделирования сложных организационно-технических систем (СОТС), к числу которых относятся и ИС, показал, что реализацию концепции рассматриваемого моделирования целесообразно осуществлять в рамках интегрированной системы поддержки принятия решений (ИСППР). В состав указанной системы в общем случае должны входить [14,17]: имитационная система, расчётно-логическая система, интеллектуальные пакеты прикладных программ, экспертная система, инструментальные CASE-средства автоматизации проектирования.

Представленную формулой (1) проблему многокритериального выбора эффективных стратегий (программ) формирования ЕИП (ЕИИП) на основе технологий, разрабатываемых в проекте Индустрия 4.0. (в частности технологий КФС),предлагается решить в рамках следующей многоэтапной процедуры, в соответствии с которой на первой фазе должно осуществляться формирование (генерирование) допустимых вариантов многоструктурных макросостояний распределенного комплекса КФС или, говоря другими словами, должен проводиться структурно-функциональный синтез нового облика ЕИП (ЕИИП) предприятия, соответствующего складывающейся (прогнозируемой) обстановке. На данном этапе на основе многомерного ортогонального проектирования на множество достижимости, сформированного в результате комплексного моделирования структурной динамики КФС, множества, с помощью которого задаются требования, предъявляемые к частным показателям эффективности ЕИП (ЕИИП), происходит формирование множества неокончательных решений (множества недоминируемых, эффективных альтернатив, множества В. Парето). Окончательный выбор эффективных системотехнических решений должен осуществляться из указанного множества. На второй фазе проводится выбор и реализация конкретного варианта многоструктурного макросостояния КФС с одновременным синтезом (построением) адаптивных планов (программ) управления переходом КФС из текущего в требуемое (выбранное) макросостояние. При этом рассматриваемые планы должны обеспечивать такое эволюционное развитие ЕИП (ЕИИП), при котором наряду с реализацией программ перехода из соответствующих макросостояний предусматривается одновременно и реализация программ устойчивого управления КФС в промежуточных макросостояниях.

Говоря о программно-техническом уровне реализации предложенной обобщенной процедуры, целесообразно в этом случае ориентироваться на сервис-ориентированную архитектуру (СОА). С понятием SOA в практическом применении тесно связано понятие сервисной шины предприятия (ESB, enterprise service bus). Сервисная шина и, прежде всего, её модельно-алгоритмическое обеспечение, по своей сути, обеспечивает автоматическую координацию работ распределённых и взаимосвязанных кибер-физических систем, реализованных в виде структуры Индустриального интернета [17]. Она обеспечивает аппаратно-программное взаимодействие разнородных киберфизических устройств, фактически управляя производством. В

общем случае сервисная шина реализует событийно-ориентированную архитектуру, то есть обеспечивает реакцию на определённые события в системе.

Применения предлагаемых технологий для использования при выполнении работ на радиационно опасных производствах

В докладе показано, как предлагаемые технологии могут быть использованы для производства работ на радиационноопасных предприятиях. Для указанных предприятий характерны следующие особенности:

1. Имеющаяся технология радиационного контроля на них нуждается в модернизации и современных подходах, при этом основные элементы ситуационно-кризисных центров и средств радиационного контроля уже развернуты на всех таких предприятиях.
2. Опасные производственные факторы радиационноопасных производств, как правило, могут быть обнаружены только техническими средствами и не определяются органами чувств человека.
3. «Носителями» веществ, представляющие радиационное загрязнение местности, производственных территорий и помещений, оборудования, механизмов и средств транспортировки, могут быть пыль, аэрозоли и технологические и рабочие жидкости, загрязненные инструменты, спецодежды и спецобуви работников, детали различных механизмов и части трубопроводов и их арматуры. Как правило, распространение радиационного загрязнения проходит в случайном порядке, и предсказать пути его распространения невозможно.

Рассматриваемые в докладе распределенные системы измерений и вычислений вместе с системами сбора данных от таких датчиков и анализа радиационных полей могут позволить выйти на новый уровень радиационной безопасности. Современные подходы к измерению опасных производственных факторов могут позволить получитьследующие комплексные информационные характеристики в реальном масштабе времени:

1. Временные, включая скорость изменения мощности эквивалентной дозы гамма излучения во времени.
2. Пространственные, включая скорость изменения мощности эквивалентной дозы гамма-излучения в пространстве, по выбранным направлениям и зонам радиационного поля.
3. Системотехнические комплексные характеристики радиационных полей:

- мощность дозы гамма-излучения от оборудования и строительных конструкций и ее пространственное распределение; поверхностная загрязненность оборудования и строительных конструкций альфа-, бета-, гамма-излучающими радионуклидами;
- радионуклидный состав источников остаточной активности в зданиях, оборудовании и на площадке; суммарная и удельная наведенная активность и активность коррозионных отложений в трубопроводах и оборудовании; поверхностное и глубинное распределение активности в материалах и конструкциях биологической защиты за счет облучения нейтронами и радиоактивного загрязнения;
- уровни загрязненности коммуникаций, строительных и защитных конструкций хранилищ продуктами деления и другими радионуклидами;
- мощности доз гамма-излучения вне зданий и сооружений блока АС;
- распределение поверхностной и глубинной активности радионуклидов на площадке;
- аэрозольная активность в воздухе помещений и боксов;

При формировании ЕИП (ЕИИП) появляются новые информационно-аналитические возможности динамического радиационного контроля, которые позволят:

1. Обеспечить персонал радиационно опасных производств и локальные и глобальные ситуационно-кризисные центры актуальной и своевременной информацией о радиационных полях и их динамике.
2. Предоставить полную и достоверную информацию федеральным и местным органам власти для своевременного принятия решений о защите населения средствами ГО и РСЧС.

В качестве конкретного примера в докладе обоснована целесообразность комбинированного подхода к повышению достоверности и точности определения состояния и местоположения носимых/мобильных комплектов сбора данных о производственной и окружающей обстановке

при их групповом управлении за счет использования дополнительных каналов, обеспечивающих одновременное решение как задач мониторинга их состояния, так и задач позиционирования на основе комбинированного использования технологий RFID, Zig-Bee и GPS/Глонасс (на открытых промплощадках).

Различные виды датчиков могут вносить различные данные в каждый конкретный момент времени и точке пространства. Например, в области данных о фактической радиационной обстановке это могут быть газоразрядные счетчики, полупроводниковые и сцинтилляционные детекторы и датчики гамма-квантов. Обладая высокой достоверностью, но из-за различных путей прохождения гамма-квантов через датчики и различного пространственного положения датчиков, они могут выдавать различные показания. Это может быть обусловлено рядом причин:

- неравномерность распределения источников радионуклидов, чаще всего имеющих случайный характер распределения в трехмерном пространстве;
- одновременное сочетанное воздействие внутренних и внешних природных и техногенных факторов радиационного характера;
- сложный энергетический спектр излучающих радионуклидов с разными показателями линейной потери энергии (ЛПЭ).

Исследования показывают, что взвешенное среднее для совокупности методов при оценке координат и собранных в данных координатах данных будет более точной оценкой, чем измерения, проведенные независимопо отдельным каналам.

Если способов оценки координат более чем два, то можно итерационно строить последовательность взвешенных сумм. Используя в качестве одного из слагаемых результат, полученный на предыдущем шаге. Если методов определения координат много, то объединять их следует не последовательно, а по типу бинарного дерева, уменьшая общее число требуемых операций взвешивания. В рамках создания методики оценивания эффективности и целесообразности совместного использования систем радиочастотной идентификации (RFID-систем) и ZigBee/LoraWan технологии были предложены несколько методов расчёта различных количественных характеристик внедряемых систем позиционирования, мониторинга и передачи данных от мобильных комплектов датчиков: оценка пропускной способности, оценка распределения нагрузок, оценка стоимости внедрения.

Заключение

Одной из важнейших проблем, которую придется решать при цифровой трансформации экономики РФ, является проблема интеграции данных и информации, поступающей из различных источников, в том числе и отКФС, число которых начинает возрастать огромными темпами. В статье предлагается методология и технология такой интеграции, базирующаяся на формировании соответствующих единых информационных и интеллектуальных пространств. В докладе приведены и формально описаны основные элементы указанных пространств и отношения между ними. В качестве примера использования предлагаемого подхода рассматривается задача формирования ЕИП при выполнении работ на радиационно опасных производствах

Исследования, выполненные по данной тематике, проводились при частичной финансовой поддержке грантов РФФИ (№№ 16-07-00779, 16-08-00510, 16-08-01277, 16-29-09482-офи-м, 17-08-00797, 17-06-00108, 17-01-00139, 17-20-01214, 17-29-07073- офи-м, 18-07-01272, 18-08-01505), в рамках бюджетной темы №№0073–2018–0003.

ЛИТЕРАТУРА:

1. Beilei Sun Xi Li, Bo Wan, Chao Wang, Xuehai Zhou, Xianglan Chen. Definitions of Predictability for Cyber Physical Systems.*Journal of Systems Architecture*. 2016. DOI: 10.1016/j.sysarc.2016.01.007.
2. Thiele L. and R. Wilhelm. Design for Timing Predictability.*Real-Time Systems*. 2004. #28. P.157–177. DOI:10.1007/11561163_14
3. Merlino Giovanni, Arkoulis Stamatis, Distefano Salvatore, Papagianni Chrysa, Puliafito Antonio, Papavassiliou Symeon. Mobile crowdsensing as a service: A platform for applications on top of sensing Clouds.*Future Generation Computer Systems*. 2016. #56. P.623–639. DOI:10.1016/j.future.2015.09.017

4. S. Distefano, G. Merlino, A. Puliafito. Sensing and actuation as a service: A new development for clouds.*Proceedings of the 2012 IEEE 11th International Symposium on Network Computing and Applications, NCA '12, IEEE Computer Society, Washington, DC, USA, 2012*. P.272–275.
5. Ganti R., Ye F., Lei H. Mobile crowdsensing: current state and future challenges.*IEEE Communications Magazine*. 2011. #49(11). P.32–39. DOI:10.1007/978–3–319–26401–1_25
6. Fei Hua, Yu Lua, Athanasios V. Vasilakos, Qi Haoc, Rui Maa, Yogendra Patil, Ting Zhanga, Jiang Lua, Xin Li, Neal N. Xiong. Robust Cyber–Physical Systems: Concept, models, and implementation.*Future Generation Computer Systems*. 2016. Vol.56. P.449–475. DOI:10.1016/j.future.2015.06.006
7. Hahn A., Ashok A., Sridhar S., Govindarasu M. Cyber-physical security testbeds: Architecture, application, and evaluation for smart grid.*IEEE Transaction of Smart Grid*. 2013. #4(2). P.847–855. DOI:10.1109/TSG.2012.2226919
8. Byung-Cheol Mina,Yongho Kima, Sangjun Leea, Jin-Woo Jungb, Eric T. Matsona. Finding the optimal location and allocation of relay robots for building a rapid end-to-end wireless communication.*Ad Hoc Networks*. 2016. Vol.39. P.23–44. DOI:10.1016/j.adhoc.2015.12.001
9. Liangyi Gonga,Wu Yanga, Zimu Zhoub, Dapeng Mana, Haibin Caic, Xiancun Zhoud, Zheng Yange. An adaptive wireless passive human detection via fine-grained physical layer information. *Ad Hoc Networks*. Vol. 38. 2016. pp. 38–50. DOI:10.1016/j.adhoc.2015.09.005
10. Kosba A.E., Saeed A., Youssef M. RASID: a robust WLAN device-free passive motion detection system.*Proceedings of IEEE International Conference on Pervasive Computing and Communications (PerCom), 2012*. P.180–189.
11. Xiao J., Wu K., Yi Y., Wang L., Ni L. FIMD: fine-grained device-free motion detection.*Proceedings of IEEE International Conference on Parallel and Distributed Systems (ICPADS), 2012.* P.229–235.
12. Kiran Joshi, Dinesh Bharadia, Manikanta Kotaru, Sachin Katti. WiDeo: Fine-grained Device-free Motion Tracing using RF Backscatter. *Proc. of 12th USENIX Symposium on Networked Systems Design and Implementation (NSDI'15), 2012*.P. 189–204.
13. Yasin Kabalci. A survey on smart metering and smart grid communication.*Renewable and Sustainable Energy Reviews.*2016. Vol.57. P.302–318. DOI:10.1016/j.rser.2015.12.114
14. Охтилев М.Ю., Соколов Б.В., Юсупов Р.М.Интеллектуальныетехнологиимониторингаиуправленияструктурнойдинамикойсложныхтехнических объектов. М.: Наука, 2006. 410 с.
15. Соколов Б.В., Юсупов Р.М. Комплексное моделирование функционирования автоматизированной системы управления навигационными космическими аппаратами.*Проблемы управления и информатики*. 2002. №5. С.1–15.
16. Соколов Б.В., Юсупов Р.М. Концептуальные основы оценивания и анализа качества моделей и полимодельных комплексов.*Теория и системы управления*. 2004. №6. С.5–16.
17. http://litsam.ru
18. Boris Sokolov, Alexandre Dolgui, Dmitry Ivanov. Optimal Control Algorithms and Their Analysis for Short-Term Scheduling in Manufacturing Systems. *Algorithms*. 2018. 11. 57. doi:10.3390/a11050057
19. Alexandre Dolgui, Dmitry Ivanov & Boris Sokolov. Ripple effect in the supply chain: an analysis and recent literature. *International Journal of Production Research*. 2018. 56:1-2.P. 414–430. DOI: 10.1080/00207543.2017.1387680.
20. Inna Trofimova, Boris Sokolov, and Dmitry Ivanov. Multiple-Model Description and Control Construction Algorithm of Supply Chain.*Cybernetics and Algorithms in Intelligent Systems. Proceedings of 7th Computer Science On-line Conference 2018*. Volume 3. P.102–108. ISSN 2194-5357 ISSN 2194-5365 (electronic)
21. Р.М. Юсупов, Б.В. Соколов, А.И. Птушкин, А.В. Иконникова, С.А. Потрясаев, Е.Г. Цивирко. Анализ состояния исследований проблем управления жизненным циклом искусственно созданных объектов. Труды СПИИРАН. 2011. №1(16). С. 37–109.

V.V. Alferov, V.V. Kondrat'ev, B.V.Sokolov, M.Y. Okhtilev, V.I. Salukhov (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences)

The report proposes a conceptual and formal model of a single intellectual information space (SIIS) based on a new approach to the distribution of sensory, network, computing, information-management and service tasks between devices built into the subject's clothes, mobile client devices, stationary service equipment, cloud computing and information resources. The problem of structurally-functionally-parametric synthesis of the distributed cyberphysical system and the corresponding SIIS is formalized, the possibility of its solution by methods of knowledge engineering, ontology theory, control theory of complex logical-dynamic objects with a rearranged structure, and methods developed in the study of operations and the theory of pattern recognition is presented. The report is considered as an example where it is necessary to implement the proposed approach to the formation of the SIIS, the task of integrating data and information on the state of facilities at radiation hazardous enterprises.

МЕТОД АДАПТАЦИИ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ ДЛЯ ДИНАМИЧЕСКИХ СИСТЕМ

В. П. Иванов
(*Федеральное государственное бюджетное учреждение науки*
«Санкт-Петербургский институт информатики и автоматизации
Российской академии наук»,
Санкт-Петербург, 199178, В.О., 14-я линия, д.39,
vpivanov.spb.su@gmail.com)

Аннотация

Рассматривается терминальная задача оптимального управления динамическими системами. Доказываются, что оптимальная траектория динамических систем является огибающей семейства сингулярных (в общем случае) кривых, названных мгновенными решениями, построенных из каждой её точки, и что оптимальное управление можно найти на указанном семействе. Таким образом, оптимальная траектория динамических систем представляется как огибающая семейства мгновенных решений. Показано, что оптимальное управление может быть найдено на семействе мгновенных решений. Предложен конструктивный метод построения мгновенных решений. Найденное таким образом управление обладает адаптивными свойствами по отношению к точности представления математической модели динамической системы, возмущениям, вычислительной устойчивости решения краевой задачи. Рассмотрен иллюстрирующий пример.

Ключевые слова: динамические системы, оптимальное управление, адаптация.

Введение

Предлагаемый доклад посвящён методу адаптации оптимального управления. Известно, что математическая модель, по которой строится оптимальное управление, не всегда адекватна динамической системе. Существуют известные проблемы решения краевой задачи (точность, вычислительная устойчивость, особенно для нелинейных динамических систем). Трудности преодолеваются рядом способов, в том числе построением адаптивного оптимального управления. К настоящему времени разработан и применяется ряд методов, таких, как теория дуального управления [1], стохастической аппроксимации [2] и др. (см. [3-5]). Актуальность проблемы вынуждает разрабатывать другие подходы её решения на иной методологической основе. В докладе рассматривается один из возможных путей решения проблемы – метод, основанный на представлении оптимальной траектории как огибающей параметрического семейства мгновенных решений.

1. Постановка задачи

Рассмотрим автономную динамическую систему вида:

$$\begin{aligned} &\frac{dx_j}{dt} = f_j(x) + B_j(x)u_j, \quad j = 1, \dots m, \\ &\frac{dx_i}{dt} = f_i(x), \qquad\qquad i = m+1, \dots n, \end{aligned} \quad , \tag{1}$$

где: $f = (f_1, \dots, f_n)$ и $B = (B_1, \dots, B_m)$ - заданные нелинейные вектор-функции; $B_j \neq 0$, $j = 1, \dots, m$; $u = (u_1, \dots, u_m)$ - m-мерный вектор управления; $m < n$, $u \in U$; U - заданное множество допустимых управлений; t - действительная переменная (время); $t \in \Im(t)$; $\Im(t)$ - открытое множество вещественной оси t; $\Im(t) = (-\infty, \dots, +\infty)$; $x = (x_1, \dots, x_n)$ - вектор состояния действительного n-мерного пространства. Все векторы принимают лишь вещественные значения.

Если система уравнений (1) неавтономная, то она приводится к виду (1) соответствующим введением дополнительной переменной.

В начальный момент времени t_0, $t_0 \in \Im(t)$ известны начальные значения вектора $x(t_0) = x_0$.

В рамках данной статьи будем рассматривать случай, когда каждая j-тая компонента управления ограничена отрезками $\left[U^{dop}_{j\,min},\ U^{dop}_{j\,max}\right]$.

Задан терминальный функционал (задача Майера):

$$J = F\left[x_i(T),\quad i = m+1,\ldots,n\right], \tag{2}$$

определенный на решениях системы уравнений (1). Где F - некоторая функция; $T \in \Im(t)$. Задача Майера была принята за основную по причине однородности структуры и возможности стандартизации математических построений. Отметим, что задачи Лагранжа и Больца могут быть сведены к задаче Майера при введении соответствующих дополнительных переменных.
Так как система уравнений (1) автономная, то множество $\Im(t)$ допустимо сузить до отрезка $\left[t_0, T\right]$. Момент времени T не фиксирован. В момент $t = T$ могут быть заданы дополнительные условия:

$$h_i = h_i\left[x(T)\right],\quad i = 1,\ldots,n, \tag{3}$$

которые могут быть включены в функционал (2) через дополнительные множители Лагранжа.

Сформулируем задачу оптимального управления следующим образом: среди всех допустимых на отрезке $\left[t_0, T\right]$ управлений $u \in U$, переводящих точку (t_0, x_0) в точку $(T, x(T))$, найти такие, для которых функционал (2), определенный на решениях системы уравнений (1), принимает наименьшее значение при выполнении условий (3).

Введем вектор-функцию множителей Лагранжа $p = (p_1,\ldots,p_n)$ и составим гамильтониан задачи оптимизации H:

$$H = \sum_{i=1}^{n} p_i f_i + \sum_{j=1}^{m} p_j B_j u_j\ . \tag{4}$$

С использованием функции H в пространстве переменных $D^n(x,p)$, $x \in D^n(x,p)$, $p \in D^n(x,p)$, сопряжённые уравнения для x и p запишутся в следующей канонической форме:

$$\begin{aligned} \frac{dx}{dt} &= \frac{\partial H}{\partial p}, \\ \frac{dp}{dt} &= -\frac{\partial H}{\partial x}. \end{aligned} \tag{5}$$

Отметим, что H и p на оптимальном решении непрерывны и к этому же приводит аналог условия Эрдмана–Вейерштрасса классического вариационного исчисления. Непрерывность сохраняется и в том случае, когда правые части уравнений (1) терпят разрыв.

Для оптимального управления $u(t)$ и фазовой траектории $x(t)$ в рамках принципа максимума необходимо существование такого ненулевого вектора p, что выполняются следующие условия:

1) Функция H переменного $u \in U$ при каждом $t \in \left[t_0, T\right]$, т.е. при фиксированных x, p, достигает при $u = u_{opt}(t)$ минимума:

$$H(x_{opt}, u_{opt}, p) = \min_{u \in U} H(x, u, p)\ . \tag{6}$$

Таким образом, оптимальное управление на границе множества допустимого управления определяется как:

$$u_{opt} = arg\ \min_{u \in U} H(x, u, p)\ . \tag{7}$$

2) Выполняются условия трансверсальности:

$$\left[H\,\delta t - \sum_{i=1}^{n} p_i\,\delta x_i\right]_{t_0}^{T} + \left[\sum_{i=1}^{n}\left(\mu_i \frac{\partial h_i}{\partial x_i} + \frac{\partial F}{\partial x_i}\right)\delta x_i\right]_{t_0}^{T} = 0\,, \tag{8}$$

где δt, δx_i - произвольные вариации соответствующих переменных; $\mu = (\mu_1, \ldots, \mu_n)$ — вектор констант.

Обобщенные условия трансверсальности в силу независимости вариаций приводят к соотношениям:

$$\begin{aligned} &\left[H\right]_{t_0}^{T} = 0, \\ &p_i = \left[\frac{\partial F}{\partial x_i} + \mu_i \frac{\partial h_i}{\partial x_i}\right], \quad i = 1, \ldots, n. \end{aligned} \tag{9}$$

Непосредственным следствием системы уравнений (5) и условия (6) является выполнение соотношения: $\frac{dH}{dt} = \frac{\partial H}{\partial t}$.

С учетом (9) для автономных систем при незаданном явно аргументе имеем:

$$H = const = 0. \tag{10}$$

Из соотношения (7) ввиду особой структуры уравнений динамической системы (1) и, соответственно, гамильтониана (4), оптимальное управление определяется как

$$u_{opt} = arg \min_{u \in U} H(x, u, p) = arg \min_{u \in U} \left(\sum_{i=1}^{n} p_i f_i + \sum_{j=1}^{m} p_j B_j u_j \right) = arg \min_{u \in U} \left(\sum_{j=1}^{m} p_j B_j u_j \right). \tag{11}$$

Откуда после преобразований имеем:

$$u_{opt\ j} = \begin{cases} U_{j\ min}^{dop}, & если \ \ p_j B_j > 0, \\ u_{j\ osob}, & если \ \ p_j = 0, \\ U_{j\ max}^{dop}, & если \ \ p_j B_j < 0, \end{cases} \qquad j = 1, \ldots, m \tag{12}$$

где $u_{j\ osob}$ — особое оптимальное управление.

Отметим, что если в начальный момент времени значения x известны (или могут быть оценены), то вектор p определен (с точностью до констант) лишь на правом конце фазовой траектории. Возникает специфическая краевая задача, после решения которой («в принципе») тем или иным способом можно найти $p(t)$, а, следовательно, и u_{opt}.

Однако, вычислительные трудности, стоящие на этом пути, методические ошибки численных методов и ошибки округлений делают процесс нахождения достоверных значений весьма трудным, а зачастую (например, при выполнении требования реального масштаба времени) и невозможным. Поэтому представляется желательным использовать нетрадиционные методы синтеза оптимального управления, одним из которых, в частности, является метод огибающих.

2. Основной результат

Сначала докажем, что оптимальная траектория динамической системы в фазовом пространстве, определяемая в смысле минимизации функционала (2), является огибающей семейства мгновенных решений, проведенных из каждой ее точки, и что оптимальное управление может быть найдено на семействе мгновенных решений.

Представим уравнение (10) в следующем виде:

$$H\left(x, u_{opt}(x, p), p\right) = H(x, p) = 0 \tag{13}.$$

Введем непрерывную функцию $W(x)$ такую, что $W(x(T)) = J = F\left[x(T)\right]$ и $p = \frac{\partial W}{\partial x}$, и приведем уравнение (13) к уравнению Гамильтона-Якоби:

$$H(x, p) = H\left(x, \frac{\partial W}{\partial x}\right) = 0. \tag{14}$$

Так как функция W входит в уравнение (14) только своими частными производными, то она определяется с точностью до аддитивной постоянной W_0, т.е.: $W = W_p(x, \alpha) + W_0$, где $\alpha = (\alpha_1, \ldots, \alpha_n)$ - вектор независимых параметров, а $W_p(x, \alpha)$ - решение уравнения (14).

Запишем уравнение гиперповерхности

$$S = W - W_p(x, \alpha) - W_0 \tag{15}$$

в $(n+1)$-мерном пространстве переменных $W, x_1, \ldots, x_n$. Из-за наличия $(n+1)$ независимых параметров $W_0, \alpha_1, \ldots, \alpha_n$ эта гиперповерхность является $(n+1)$-параметрическим семейством гиперповерхностей. Параметр W_0 вызывает лишь сдвиг вдоль оси W, поскольку $\frac{\partial S}{\partial W_0} = -1 \neq 0$. Таким образом, огибающих относительно параметров $W_0, \alpha_1, \ldots, \alpha_n$ нет. Для того чтобы построить огибающие, уменьшим размерность параметров и примем:

$$\begin{aligned} W_0 &= W_0(\gamma_1, \ldots, \gamma_n) \\ \alpha_\nu &= \alpha_\nu(\gamma_1, \ldots, \gamma_n), \quad \nu = 1, \ldots, n \end{aligned} \tag{16}$$

где $\gamma_1, \ldots, \gamma_n$ - параметры, такие, что $\frac{\partial W}{\partial \gamma_i} \neq 0, \quad \frac{\partial \alpha_\nu}{\partial \gamma_i} \neq 0, \quad i = 1, \ldots, n, \quad \nu = 1, \ldots, n$.

Построим огибающие относительно параметров $\gamma_1, \ldots, \gamma_n$:

$$\frac{\partial S}{\partial \gamma_\nu} = \sum_{i=1}^{n} \frac{\partial W}{\partial \alpha_i} \cdot \frac{\partial \alpha_i}{\partial \gamma_\nu} - \sum \frac{\partial W_p}{\partial \alpha_i} \cdot \frac{\partial \alpha_i}{\partial \gamma_\nu} - \frac{\partial W_0}{\partial \gamma_\nu} = 0, \quad \nu = 1, \ldots, n. \tag{17}$$

Выбором функций (16) добьемся, чтобы $det\left(\frac{\partial \alpha_i}{\partial \gamma_\nu}\right) \neq 0, \quad i, \nu = 1, \ldots, n$.

Тогда уравнения (17) будут иметь решения:

$$\beta_\nu = \frac{\partial W}{\partial \alpha_\nu}, \quad \nu = 1, \ldots, n. \tag{18}$$

С другой стороны, если W - полный интеграл уравнения (14), то по теореме Якоби имеем:

$$\beta = \frac{\partial W}{\partial \alpha}, \qquad p = \frac{\partial W}{\partial x}. \tag{19}$$

Потребуем, чтобы α и β удовлетворяли преобразованию гамильтониана $H(x, p)$ в гамильтониан $H(\alpha)$, а также каноническим уравнениям, которые, ввиду (14), запишутся как

$$\frac{d\alpha}{dt} = -\frac{\partial H}{\partial \beta} = 0, \quad \frac{d\beta}{dt} = \frac{\partial H}{\partial \alpha} = 0. \tag{20}$$

Из (16-20) следует, что оптимальные траектории являются огибающими n-параметрического семейства поверхностей.

По теореме Якоби $\beta = \frac{\partial W}{\partial \alpha}, \qquad p = \frac{\partial W}{\partial x}$. Но $W(x, \alpha)$. Тогда вектор $p = p(x, \alpha)$. То есть, p, а, следовательно, и оптимальное управление можно найти на семействе поверхностей.

Отметим, что на семейство поверхностей не налагается никаких ограничений, кроме условий огибания их оптимальной траекторией. Поэтому возможны разнообразные подходы к их формированию и, соответственно, к синтезу управления.

В рамках данной работы предлагается ещё один подход к формированию огибания.

Представим функцию $W_p(x, \alpha)$ в виде суммы функций, каждая из которых зависит только от одной из переменных $x_\nu, \nu = 1, \ldots, n$, т.е.:

$$W = \sum_{\nu=1}^{n} W_\nu^{\kappa}(x_\nu, \alpha) + W_0. \tag{21}$$

Определим канонические переменные p_ν, $\nu = 1, \ldots, n$:

$$p_\nu = \frac{\partial W}{\partial x_\nu} = \frac{\partial W_\nu^\kappa(x_\nu, \alpha)}{\partial x_\nu}, \quad \nu = 1, \ldots, n. \tag{22}$$

Согласно выражениям (22), переменные p_ν, $\nu = 1, \ldots, n$ оказываются функциями только одной x_ν и α, в то время, как уравнения (1), (5), (19) говорят о том, что p_ν, $\nu = 1, \ldots, n$ в общем случае должны быть функциями всех $x_1, \ldots, x_n$ и остальных p_i, $i = 1, \ldots, n$, $i \neq \nu$. Это противоречие может быть устранено, если приравнять α_ν, $\nu = 1, \ldots, n$ некоторые определенные комбинации переменных $x_1, \ldots, x_n$, «замороженных» в данный момент времени, т.е.:

$$\alpha_\nu = \alpha_\nu(x_1, \ldots, x_n), \quad \nu = 1, \ldots, n. \tag{23}$$

Из выражений (16), (19), (22), (23) следует, что $p_1, \ldots, p_n$ и управление u_{opt}, можно определить на параметрическом семействе поверхностей, которое огибает оптимальная траектория, если в качестве параметров $\gamma_1, \ldots, \gamma_n$ соответствующим образом взять фазовые координаты, а уравнения (17) будут разрешимы относительно частных производных W_p по параметрам $\alpha_1, \ldots, \alpha_n$. Фиксируя в качестве параметров «замороженные» в текущий момент времени значения фазовых координат, мы тем самым на семействе поверхностей определим семейство кривых. Назовем их мгновенными решениями, поскольку они определяются функцией W_p, являющейся решением уравнения (14).

В рамках решения уравнения (14) можно использовать в качестве мгновенных решений произвольные заданные функции, удовлетворяющие условиям (17), (19) - (23).

С другой стороны, соотношения (18)-(23) говорят о том, что для построения мгновенных решений можно производить условное разделение переменных в уравнении Гамильтона-Якоби (см. [4]).

Отметим, что мгновенные решения должны удовлетворять условию минимизации функционала (2) относительно используемых параметров.

Таким образом, оптимальная траектория динамической системы в фазовом пространстве является огибающей семейства мгновенных решений, проведенных из каждой ее точки.

Представим уравнения (5) для $p = (p_1, \ldots, p_n)$ в следующей форме:

$$\frac{dp_\nu}{dt} = -\frac{\partial H}{\partial x_\nu} = -p_\nu\left(\frac{\partial f_\nu}{\partial x_\nu} + \frac{\partial B_\nu}{\partial x_\nu}\cdot u_\nu\right) - \sum_{\substack{j=1 \\ j\neq\nu}}^{m} p_j\left(\frac{\partial f_j}{\partial x_\nu} + \frac{\partial B_j}{\partial x_\nu}\cdot u_j\right) - \sum_{i=m+1}^{n} p_i \frac{\partial f_i}{\partial x_\nu} \tag{24}$$

$\nu = 1, \ldots, m$

Преобразуем это уравнение к виду:

$$\frac{dp_\nu}{dt} + \Phi_\nu\, p_\nu = G_\nu \qquad \nu = 1, \ldots, m, \tag{25}$$

где: $\Phi_\nu = \left(\frac{\partial f_\nu}{\partial x_\nu} + \frac{\partial B_\nu}{\partial x_\nu}\cdot u_\nu\right)$, $G_\nu = -\sum_{\substack{j=1 \\ j\neq\nu}}^{m} p_j\left(\frac{\partial f_j}{\partial x_\nu} + \frac{\partial B_j}{\partial x_\nu}\cdot u_j\right) - \sum_{i=m+1}^{n} p_i \frac{\partial f_i}{\partial x_\nu}$.

Проинтегрировав уравнение (25), получим:

$$p_\nu = exp(-\int \Phi_\nu\, dt)\left[\int G_\nu\, exp(\int \Phi_\nu\, dt) - C_\nu\right] \quad \nu = 1, \ldots, m. \tag{26}$$

Постоянные C_ν, $\nu = 1, \ldots, n$ находятся из условий трансверсальности (9).

Отметим, что знак функций p_ν, $\nu = 1, \ldots, n$ определяется знаком функций G_ν.

В работе [3] доказано, что ввиду линейного вхождения управления в систему уравнений (1), особое оптимальное управление каждой j-той компоненты может быть найдено из системы уравнений:

$$\frac{d^k}{dt^k}\left(\frac{\partial H}{\partial u_j}\right)=0, \qquad k=0,1,\ldots,2p_s \tag{27}$$

(p_s - порядок сингулярности) при выполнении следующих необходимых условий оптимальности:

$$(-1)^{p_s}\frac{\partial}{\partial u_j}\left[\frac{d^{2p_s}}{dt^{2p_s}}\left(\frac{\partial H}{\partial u_j}\right)\right]\geq 0, \qquad p_s=0,1,2,\ldots \tag{28}$$

если $det\left\{\dfrac{\partial^2 H}{\partial u_i\,\partial u_j}\right\}\equiv 0, \quad i,j=1,\ldots,m.$

Первоначально рассмотрим случай, когда порядок сингулярности равен единице. Тогда особое управление можно найти из системы уравнений:

$$p_j=0, \quad \frac{dp_j}{dt}=0, \quad \frac{d^2p_j}{dt^2}=0, \qquad j=1,\ldots,m. \tag{29}$$

Из первых двух уравнений системы (29) следует, что $G_j=0$. Разрешим это уравнение относительно переменной x_j. Если корень существует, то:

$$x_j=\eta_j(x_\nu, p_\nu;\ \nu=1,\ldots,n;\ \nu\neq j). \tag{30}$$

Тогда третье уравнение системы (29) после преобразований запишем как:

$$\frac{dG_j}{dt}=\frac{\partial G_j}{\partial x_j}\left(f_j+B_j u_{j\,osob}\right)+\frac{\partial G_j}{\partial \eta_j}\cdot\frac{d\eta_j}{dt}=0.$$

Из последнего соотношения можно найти особое управление:

$$u_{j\,osob}=-\frac{f_j+\left(\dfrac{\partial G_j}{\partial \eta_j}\cdot\dfrac{d\eta_j}{dt}\right)\Big/\dfrac{\partial G_j}{\partial x_j}}{B_j} \tag{31}$$

при выполнении необходимых условий оптимальности в следующей форме:

$$B_j\left.\frac{\partial G_j}{\partial x_j}\right|_{x_j=\eta_j}\leq 0, \qquad j=1,\ldots,m. \tag{32}$$

Отметим, что корень уравнения $G_j=0$ необязательно может быть единственным. Тогда каждый корень проверяется на выполнение необходимых (32) и достаточных (2) условий оптимальности.

Отметим также, что на мгновенном решении $x_j=const=\eta_j$.

Из (26), (29), (30) также следует, что на мгновенном решении

$$sign\ p_j=-sign(x_j-\eta_j).$$

В случае нулевого порядка сингулярности p_j (см. (26)) либо сохраняет знак, либо становится равным нулю лишь в одной точке на конце отрезка. Дополнив условие $p_j=0$ условиями первого порядка сингулярности, мы приходим к аналогичному выражению.

В случае, когда порядок сингулярности превышает единицу, то, выделив условие сингулярности первого порядка $G_j=0$, или, что тоже самое, $x_j=\eta_j$, а затем, продифференцировав его $2p_s-1$ раз, можно найти особое управление как:

$$u_{j\,osob} = f\left(x, \eta_j, \frac{d\eta_j}{dt}, \ldots, \frac{d^{2p_s-1}\eta_j}{dt^{2p_s-1}} \right). \tag{33}$$

С учетом полученных выражений управление (12) примет вид:

$$u_{opt\ j} = \begin{cases} U_{j\,min}^{dop}, & если\ \ (x_j - \eta_j) B_j < 0, \\ u_{j\,osob}, & если\ \ x_j = \eta_j, \\ U_{j\,max}^{dop}, & если\ \ (x_j - \eta_j) B_j > 0, \end{cases} \tag{34}$$

$$j = 1, \ldots, m$$

Следовательно, управление в рамках заданного терминального критерия оптимальности можно найти на семействе мгновенных решений.

Обратим внимание на то, что закон управления в выражениях (33, 34) обладает адаптивными свойствами и строится по принципу обратной связи по отношению к параметрам, определенным на семействе мгновенных решений. Адаптация осуществляется по вычисляемым в каждый момент времени параметрам на мгновенных решениях и соответствующим производным.

3. Пример

Требуется найти управление, минимизирующее функционал:

$$J = \sqrt{\left(x_3(T) - x_3^k\right)^2 + \left(x_4(T) - x_4^k\right)^2},$$

где x_i^k, $i = 3, 4$ - заданные значения соответствующих координат, определенный на решениях следующей системы уравнений:

$$\frac{dx_1}{dt} = A(u - \cos x_1),$$

$$\frac{dx_2}{dt} = B,$$

$$\frac{dx_3}{dt} = x_2 \sin x_1,$$

$$\frac{dx_4}{dt} = x_2 \cos x_1,$$

u - управление, $|u| \le U$, U ограничение на управление, A, B - положительные константы.

Составим Гамильтониан $H = \sum_{i=1}^{n} p_i f_i + \sum_{j=1}^{m} p_j B_j u_j$, а затем, с учётом того, что $H = 0$, уравнение Гамильтона-Якоби, которое примет вид:

$$\frac{\partial W}{\partial x_1} A(u - \cos x_1) + \frac{\partial W}{\partial x_2} B + \frac{\partial W}{\partial x_3} x_2 \sin x_1 + \frac{\partial W}{\partial x_4} x_2 \cos x_1 = 0.$$

Для нахождения мгновенных решений условно разделим переменные в уравнении Гамильтона-Якоби:

$$\frac{\partial W}{\partial x_1} A(u - \cos x_1) + \alpha_2 + \alpha_3 x_2 \sin x_1 + \alpha_4 x_2 \cos x_1 = 0,$$

$$\frac{\partial W}{\partial x_2} B = \alpha_2, \quad \frac{\partial W}{\partial x_3} = \alpha_3, \quad \frac{\partial W}{\partial x_4} = \alpha_4.$$

Определим функцию W :

$$W = (\alpha_2 + \alpha_3\, x_2\, sin\, x_1 + \alpha_4\, x_2 \cos x_1)(t - T) + \alpha_2 \int_{x_2(t)}^{x_2(T)} \frac{dx_2}{B} + \alpha_3 \int_{x_3(t)}^{x_3(T)} dx_3 + \alpha_4 \int_{x_4(t)}^{x_4(T)} dx_4 + W_0 .$$

Определим мгновенные решения:

$$x_1 = \eta, \quad x_2(T) = x_2(t) + B(T - t),$$
$$x_3(T) = x_3(t) + x_2\, sin\, \eta \cdot (T - t),$$
$$x_4(T) = x_4(t) + x_2\, cos\, \eta \cdot (T - t).$$

Подставляя эти значения в функционал и минимизируя последний, после преобразований получим, что $tg\,\eta = \dfrac{x_3^k - x_3(T)}{x_4^k - x_4(T)}$.

Тогда выражение для управления примет вид:

$$u = \begin{cases} -U\ sign\ sin(x_1 - \eta), & если\ x_1 \neq \eta, \\ cos\,\eta + \dfrac{1}{A} \cdot \dfrac{d\eta}{dt}, & если\ x_1 = \eta. \end{cases}$$

На мгновенных решениях выполняются соотношения: $x_1 = \eta$ или $x_1 = \eta \pm \pi$. $\eta = const$.

Для оценки сравнительной эффективности предлагаемого подхода строилась система канонических уравнений

$$\frac{dx}{dt} = \frac{\partial H}{\partial p},$$
$$\frac{dp}{dt} = -\frac{\partial H}{\partial x}.$$

находилось управление в форме (12):

$$u_{opt\ j} = \begin{cases} U_{j\ min}^{dop}, & если\ \ p_j B_j > 0, \\ u_{j\ osob}, & если\ \ p_j = 0, \\ U_{j\ max}^{dop}, & если\ \ p_j B_j < 0, \end{cases}$$
$$j = 1, \ldots, m$$

При условиях трансверсальности (9) исходная задача решалась и методом последовательных приближений. При этом x_3^k полагался равным 2325, x_4^k = 3510, U =3. Оказалось, что при близкой точности (отклонение результатов не превышало 0,3) задача методом последовательных приближений решалась за 348 с, а методом огибающих – за 17.

Для проверки адаптивных свойств добавим во второе уравнение динамической системы неучтенную ранее константу 0,12. Результаты расчета показали, что при этих условиях x_3 и x_4 в конце процесса совпадают с результатами предыдущего примера с точностью до 0,5, т.е. ошибка составляет меньше 0,1%. Полученный результат подтверждает адаптивность управления при неточном формировании математической модели.

4. Заключение

Обобщим полученные результаты. В результате проведённых исследований доказано: 1) оптимальная траектория динамических систем может быть представлена как огибающая параметрического семейства поверхностей, 2) оптимальное управление может быть найдено на параметрическом семействе поверхностей, 3) фиксируя в заданный момент времени «замороженные» значения фазовых координат, на семействе поверхностей выделяем сингулярные кривые («мгновенные» решения), которые огибает фазовая траектория, и на семействе которых может быть найдено оптимальное управление, 4) предложен конструктивный способ построения мгновенных решений, 5) управление, найденное на семействе мгновенных решений, обла-

дает адаптивными свойствами, 6) на семействе мгновенных решений в ряде случаев упрощается решение двухточечных краевых задач.

ЛИТЕРАТУРА

1. Фурасов В.Д. Устойчивость движения, оценки и сиабилизация. М.: Наука, 1977. - 248 с.
2. Цыпкин Я.З. Оптимальные адаптивные системы управления объектами с запаздыванием.// Автоматика и телемеханика, 1986. №8 - С.5.
3. Фрадков А.Л. Адаптивное управление в сложных системах: беспоисковые методы. М.: Наука, 1990. — 296 с.
4. Воронов А.А. Устойчивость, управляемость, наблюдаемость. М.: Наука, 1979. – 335 с.
5. Чураков Е.П. Оптимальные и адаптивные системы. М.: Энергоатомиздат, 1987. – 256 с.
6. Болтянский В.Г. Математические методы оптимального управления. М.: Наука, 1969. – 408 с.
7. Сейдж Э.П., Уайт Ч.С. Оптимальное управление системами. М.: Радио и связь, 1982. – 398 с.
8. Габасов Р., Кириллова Ф.М. Особое оптимальное управление. М.: Наука, 1973. – 256 с.

V.P.Ivanov (St.Petersburg institute for informatics and automation RAS, Saint-Petersburg)

Method of adaptation of optimum control for dynamic systems

The terminal problem of optimal control of dynamical systems is considered. It is proved that the optimal trajectory of dynamical systems is the envelope of a family of singular (in general) curves called instantaneous solutions constructed from each of its points, and that optimal control can be found on the specified family. Thus, the optimal trajectory of dynamical systems is represented as the envelope of the family of instantaneous solutions. It is shown that optimal control can be found on a family of instantaneous solutions. A constructive method for constructing instant solutions is proposed. The control thus obtained has adaptive properties with respect to the accuracy of the representation of the mathematical model of a dynamical system, perturbations, and the computational stability of the solution of the boundary value problem. An illustrative example is considered.

СИСТЕМНЫЙ МОНИТОРИНГ ПРИ ИНТЕЛЛЕКТУАЛЬНОЙ ПОДДЕРЖКЕ УПРАВЛЕНИЯ КРИТИЧЕСКИХ ОБЪЕКТОВ

А. В. Алексеев, Р. И. Мусатенко, И. Ю. Тычинин
(*Институт автоматизации процессов борьбы за живучесть корабля, судна, С.-Петербург, 198262, Ленинский пр., 101, оф. 162, iapbgks@bk.ru*)

Аннотация

Рассмотрены системные возможности новой технологии мониторинга состояния объектов информатизации (технологии СПРУ) на основе автоматической оценки системных показателей качества – агрегированного показателя качества, конкурентной способности и перспективности разития, а также их мониторинга, прогнозирования, интеллектуальной поддержки выработки управляющих решений, контроля и мониторинга реализации принятых управленческих решений.

Представляемая технология СПРУ обладает конкурентной способностью в сравнении с аналогичными современными, например, судовыми системами более 20%, что для условий сравнения конкурирующих объектов весьма существенно.

В сочетании с инвариантностью к специфике решаемых задач, минимальной избыточностью визуализируемых данных, гармоническим алгоритмом агрегирования данных, автоматической процедурой генерации проектов управленческих решений и выдачей их в удобной (текстовой, звуковой) форме, а также синхронной фото-, видео- и аудио-поддержкой оценки обстановки позволяет рекомендовать технологию СПРУ к самому широкому внедрению и интеграции в существующие автоматизированные рабочие места операторов систем управления безопасностью различного назначения и, прежде всего, критических объектов.

Ключевые слова: критический объект, квалиметрическое обеспечение, системный мониторинг, интеллектуальная поддержка, агрегированный показатель качества.

Введение

Обеспечение безопасности эксплуатации, локализации аварийных ситуаций и аварий, а также борьба за живучесть критических объектов сегодня характеризуется интенсивно развивающимися технологиями и системами автоматизации, роботизации и интеллектуализации управления.

Среди известных подходов это реализуется, в первую очередь, на основе комплексных систем интеллектуальной поддержки принятия решений и управления (СПРУ) с автоматической или роботизированной оценкой системных и организационно-технических показателей качества, их мониторингом, прогнозированием, интеллектуальной поддержкой обоснования принимаемых решений, контролем реализации принятых решений и упреждающим управлением [1, 2].

Именно квалиметрическое системное обеспечение этих процессов на основе количественного измерения и оценивания качества функционирования системы в целом позволяет адекватно оценивать ситуационную обстановку на объекте и принимать рациональные (безошибочные, эффективные, оптимальные) управленческие решения [3, 4], в том числе эффективно минимизировать влияние негативных субъективных свойств операторов, их «человеческий фактор», что сегодня является одной из острейших проблем управления.

Это важнейшее направление продолжает оставаться далеко не исчерпанным как в теоретическом, так и в практическом аспектах. Более того, именно системный мониторинг качества управления при интеллектуальной поддержке управления, в первую очередь, критическими объектами, как утверждает в последнее время ряд специалистов, является наиболее перспективным и востребованным, в связи с чем и рассматривается в данной работе.

Предлагаемый доклад посвящен анализу специфических особенностей, представленных в систематизированном виде в варианте СПРУ с подсистемой интеллектуальной поддержки «АПМУ.о2,» и новым возможностям мониторинга на основе непрерывного наблюдения, прогнозирования и анализа системных показателей качества объектов информатизации.

Технология СПРУ, как новое направление мониторинга качества критических объектов

Постановка задачи. Среди системных показателей именно качество, как мера соответствия объекта своему предназначению, является, как известно, основной характеристикой, отражаю-

щей его специфические свойства и возможности. Однако, определение критериев и соответствующих показателей является сложной задачей моделирования качества и эффективности, как меры реализации проектного показателя качества. В современных условиях резкого возрастания сормационном дефложности объектов анализа построение аналитических моделей качества практически становится невозможным и используется, как правило, для оценки и анализа лишь отдельных свойств объектов.

В свою очередь, именно анализ динамики качества в процессе функционирования объекта представляет наибольший интерес и, в первую очередь, с позиций системного анализа и синтеза. Поэтому поиск и освоение универсальных методов и технологий оценки и мониторинга качества является традиционно актуальной и весьма востребованной задачей.

В этом контексте и в интересах совершенствования алгоритмов и систем интеллектуальной поддержки управления, прежде всего, критическими объектами рассмотрим на конкретном примере реализации рассмотрим особенности и уникальные возможности мониторинга системных показателей качества объектов анализа.

Выбор объекта анализа. В качестве конкретного примера реализации задач мониторинга системных показателей качества в ряду с современными методами анализа и синтеза при информационном дефиците профессора Хованова Н.В. и программными средствами реализации «ASPID», «APIS», метода анализа иерархий Томаса Саати и средствами типа «MPRIORITY» рассмотрим метод и технологию систем поддержки принятия решений и управления (СПРУ), представленную в ряде работ, включая [5, 6].



Рис. 1. Назначение и основные характеристики технологии СПРУ в варианте ПИП «АМПУ.о2»

В обобщенном виде основные характеристики рассматриваемой технологии в варианте СПРУ с подсистемой информационно-аналитической и интеллектуальной поддержки принятия решений (ПИП) «АМПУ.о2» приведены на рис. 1.

Особенности технологии СПРУ. Одним из 7 важнейших конкурентных преимуществ разработанной технологии СПРУ является её инвариантность (неизменность, независимость от условий, равнопригодность) к специфике решаемых задач. Полученные результаты апробации

и внедрения позволяют рекомендовать технологию СПРУ к широкому внедрению, интеграции в существующие автоматизированные рабочие места операторов систем управления безопасностью различного назначения.

Концепция решения задач СПРУ включает количественную оценку, прежде всего, системных показателей качества на основе специального агрегирования частных (ЧПК), групповых (ГПК), модельных (МПК) в агрегированный показатель качества (АПК), их непрерывный с циклом управления от 2 минут и более мониторинг (непрерывное наблюдение), прогнозирование и контроль не только ЧПК (по показаниям разнородных датчиков технического состояния объекта наблюдения), ГПК (характеризующих соответствующие свойства объекта), но и системного, агрегированного показателя качества (АПК), характеризующего качество объекта в целом. Другие свойства и конкурентные преимущества технологии СПРУ приведены на рис. 1.

Новые возможности технологии СПРУ. В целом реализация технологии СПРУ позволяет системно целостно наблюдать и контролировать практически в автоматическом режиме АПК, как наиболее значимый системный показатель состояния критического объекта. Например, АПК в виде комбинированной гармонической свёртки групповых показателей качества типа: функциональная устойчивость; взрывопожарная и радиационная безопасность объекта; живучесть технических и специальных средств объекта; защищённость персонала объекта; устойчивость системы управления (управляемость) безопасностью, аварийных ситуаций и борьбой за живучесть.

Тем самым, мониторинг АПК критического объекта позволяет обеспечить непрерывный комплексный контроль состояния качества объекта по всей совокупности датчиков контроля, число которых может превышать 2000…3000 (частных показателей качества), а также свойств объекта с числом порядка 5…7. Причем, именно анализ динамики АПК позволяет объективно оценивать и прогнозировать обстановку на объекте и формировать своевременные рекомендации лицу, принимающему решения, по управлению этой обстановкой.

Еще более значимыми результаты системного анализа состояния критического объекта становятся при мониторинге АПК в сочетании с такими организационно-системными показателями, как конкурентная способность (КС, как отношение АПК объекта наблюдения к АПК конкурентного объекта) и перспективность развития (ПР, как отношение АПК объекта наблюдения к АПК этого же объекта, но на предыдущем этапе развития) [6 - 8].

Системные показатели КС и ПР отражают квалиметрически достигаемое при технологическом развитии объектов анализа их сравнительное превосходство, либо отставание в сравнении с принятыми за базу сравнения значениями АПК соответственно конкурентных объектов, либо вариантов объекта по отношению к предыдущему его этапу.

В свою очередь, анализ динамики системных показателей позволяет максимально информативно оценивать текущее состояние объекта анализа, результативность и эффективность принимаемых и корректирующих управленческих решений.

В целом, изложенное позволяет рассматривать представленную технологию СПРУ, как новое направление мониторинга качества, в первую очередь, критических объектов.

Заключение

Представленная технология СПРУ на основе автоматической оценки системных показателей качества (АПК, КС, ПР), их мониторинга, прогнозирования, интеллектуальной поддержки выработки управляющих решений, контроля и управления безопасностью характеризуется собственной конкурентной способностью в сравнении с аналогичными современными, например, судовыми системами более 20%, что для условий сравнения конкурирующих объектов весьма существенно [8, 9].

В сочетании с инвариантностью к специфике решаемых задач, минимальной избыточностью визуализируемых данных, гармоническим алгоритмом агрегирования данных, автоматической процедурой генерации проектов управленческих решений и выдачей их в удобной (текстовой, звуковой) форме, а также синхронной фото-, видео- и аудио-поддержкой оценки обстановки позволяет рекомендовать технологию СПРУ к самому широкому внедрению и интеграции в существующие автоматизированные рабочие места операторов систем управления безопасностью различного назначения и, прежде всего, критических объектов.

ЛИТЕРАТУРА

1. Захаров И.Г. Концептуальный анализ в военном кораблестроении. - СПб.: Судостроение, 2001.
2. Шпак В.Ф., Директоров Н.Ф., Мирошников В.И., Навойцев С.П., Наумов В.Н., Серегин А.В., Синещук Ю.И., Туровский О.М. Информационные технологии в системе управления силами ВМФ - Санкт-Петербург, "Элмор", 2005.
3. Алексеев А.В. Квалиметрическое обеспечение организации принятия проектных и управленческих решений в сложных системах / Системный анализ при создании кораблей, комплексов вооружения и военной техники. Тематический сборник. Выпуск 15. – СПб.: «Моринтех», 2008, с. 67 - 75.
4. Субетто А.И., Алексеев А.В. Теория практики квалиметрического обеспечения развития морских автоматизированных систем / Актуальные проблемы морской энергетики: материалы седьмой Всероссийской межотраслевой научно-технической конференции в рамках Второго Всероссийского научно-технического форума «Корабельная энергетика: из прошлого в будущее». – СПб.: Изд-во СПбГМТУ, 2018, с. 78 – 86.
5. Алексеев А.В., Левадный И.В., Мусатенко Р.И., Смольников А.В., Сус Г.Н., Ушакова Н.П. Технология и программный комплекс системы поддержки принятия решений и управления судоводителя в сетевом исполнении (ТиПК «СПРУ-С») / Региональная информатика (РИ-2016). Юбилейная XV Санкт-Петербургская международная конференция «Региональная информатика (РИ-2016)». Санкт-Петербург, 26-28 октября 2016 г.: Материалы конференции. \ СПОИСУ. - СПб, 2016, с. 425-426.
6. Алексеев А.В. Технология системного мониторинга и интеллектуальной поддержки управления безопасностью критических объектов - http://elibrary.ru/item.asp?id=26338682, с.215-219.
7. Алексеев А.В., Смольников А.В. Интеллектуальная поддержка принятия управленческих решений при борьбе за живучесть морских объектов повышенного риска / Материалы конференции «Информационные технологии в управлении» (ИТУ-2012). - СПб.: ОАО «Концерн «ЦНИИ «Электроприбор», 2012. С. 269-275.
8. Алексеев А.В., Антипов В.В., Бобрович В.Ю., Смольников А.В. Концепция и технология ранговой партнерской сертификации объектов морской техники и инфраструктуры / Материалы конференции «Информационные технологии в управлении» (ИТУ-2014). ГНЦ РФ ОАО «Концерн «ЦНИИ «Электроприбор». Спб., 7-9.10.2014 – СПб.: ОАО «Концерн «ЦНИИ «Электроприбор», 2014. С. 29-34.
9. Алексеев А.В., Кузнецов В.В., Равин А.А., Согонов С.А., Хруцкий О.В. Оптимизация системного управления ОМТИ: теория практики / Труды СПбГМТУ, выпуск 5 (Труды ЛКИ. Выпуск 268), 2018, с. 65 – 69.
10. Алексеев А.В., Антипов В.А., Бобрович В.Ю., Смольников А.В. Оптимизация системных решений при управлении живучестью судовых автоматизированных систем и комплексов на основе анализа их конкурентной способности / Материалы 9-й конференции «Информационные технологии в управлении» (ИТУ-2016). – СПб.: АО «Концерн «ЦНИИ «Электроприбор», 4-6.10.2016 г. - Спб., ГНЦ РФ АО «Концерн «ЦНИИ «Электроприбор», 2016.
11. Алексеев А.В., Карпов А.Е., Тычинин И.Ю. Квалиметрия защищенности и мониторинг безопасности критических объектов: методы, технологии, практика / XXI Всероссийская научно-практическая конференция «Актуальные проблемы защиты и безопасности», 3-6 апреля 2018: материалы. – СПб.: РАРАН, 2018, с. 68 – 73.

A. V. Alekseev, R. I., Musatenko, I. Y. Tychinin (Institute of automation of processes of struggle for survivability of the ship, vessel)

System monitoring when the intellectual support of the department of critical facilities Requirements to formatting the papers of the Conference

The system capabilities of the new technology for monitoring the state of Informatization objects (SPRU technology) on the basis of automatic assessment of system quality indicators – aggregate quality indicator, competitive ability and prospects of development, as well as their monitoring, forecasting, intellectual support for the development of management decisions, control and monitoring of the implementation of management decisions are considered.

The presented technology has a competitive ability in comparison with similar modern, for example, ship systems more than 20%, which is very important for the conditions of comparison of competing objects.

In combination with the invariance to the specifics of the tasks, the minimum redundancy of the visualized data, the harmonic algorithm, data aggregation, automatic procedure for the generation of projects of management decisions and their results in a convenient (textual, audio) form, and simultaneous photo-, video - and audio-supported assessment of the situation allows us to recommend technology SPRUE to the broadest possible adoption and integration into existing automated work places of operators ' safety management systems for various purposes and, especially, the critical objects.

МОДЕЛЬНО-АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ПЛАНИРОВАНИЯ РЕКОНФИГУРАЦИИ БОРТОВОЙ АППАРАТУРЫ МАЛОМАССОРАЗМЕРНЫХ КОСМИЧЕСКИХ АППАРАТОВ

А. Ю. Кулаков[1], **Ф. М. Кулаков**[2], **А. Н. Павлов**[3], **С. А. Потрясаев**[4], **Б. В. Соколов**[5]
(*Федеральное государственное бюджетное учреждение науки
«Санкт-Петербургский институт информатики и автоматизации» Российской академии наук,
199178, Санкт-Петербург, 14 линия, д. 39,
russ69@bk.ru, kufelix@yandex.ru, pavlov62@list.ru, spotryasaev@gmail.com*)

Аннотация

В докладе рассматриваются методологические, методические и технологические основы постановки и решения задач реконфигурации бортовой аппаратуры (БА) маломассоразмерных космических аппаратов (МКА). Предложено данные задачи исследовать и решать с использованием разработанных авторами прикладной теории проактивном управления структурной динамикой сложными техническими объектами. Предложено формальное описание и интерпретация решаемых задач реконфигурации БА МКА как задач динамического структурно-функционального синтеза сложных технических объектов. Разработаны соответствующие модели, методы и алгоритмы, базирующиеся на конструктивной реализации концепций комплексного моделирования и проактивного управления. Приводятся сведения об особенностях практической реализации разработанного специального модельно-алгоритмического обеспечения в рамках соответствующего экспериментального образца программного комплекса.

Ключевые слова: методы, модели и алгоритмы реконфигурации, бортовая аппаратура.

Введение

В современных условиях для обеспечения требуемой степени автономности, качества и оперативности управления состоянием такого сложного объекта как маломассоразмерных космический аппарат (МКА) и его бортовой аппаратурой (БА) необходимо решить следующие основные научно-технические проблемы. Во-первых, обеспечить модельно-алгоритмическое описание процессов смысловой интерпретации всех возможных штатных и нештатных состояний их функционирования и, во-вторых, на этой основе решить основные задачи комплексной автоматизации и интеллектуализации процессов мониторинга технического состояния и управления БА МКА в различных условиях обстановки. Однако, к сожалению, в подавляющем большинстве случаев на практике мониторинг, прогнозирование и управление состояниями элементов, подсистем и всей БА с точки зрения обеспечения требуемых уровней показателей надежности, живучести и эффективности функционирования МКА автоматизирован, в лучшем случае, лишь частично. Как правило, в современных системах мониторинга технического состояния и управления БА МКА операторам представляется смысловая информация только о состояниях их элементов, а не объектов контроля в целом. Указанные обстоятельства приводят к тому, что интегральную оценку и прогнозирование состояния БА МКА также, как и формирование необходимых управляющих воздействий выполняют операторы, в основном, вручную, базируясь эвристических правилах [1-2].

Особую актуальность вопросы автономности и живучести приобретают при разработке и эксплуатации МКА наблюдения (дистанционного зондирования Земли (ДЗЗ)) [1-2]. Для данных МКА важнейшим показателем качества их функционирования является показатель оперативности передачи и выполнения программ работы специальной и обеспечивающей БА МКА, так и получения целевой информации и телеметрии наземным комплексом управления (НКУ), а также возможности оперативного вмешательства в реализацию указанных программ в случае возникновения нештатной ситуации. Таким образом, для МКА наблюдения вопросы автономности и живучести тесно взаимосвязаны с вопросами эффективного функционирования как БА МКА, так и НКУ, и в целом АСУ МКА.

---

[1] Кандидат технических наук, младший научный сотрудник.
[2] Доктор технических наук, профессор, главный научный сотрудник.
[3] Доктор технических наук, профессор, старший научный сотрудник.
[4] Кандидат технических наук, старший научный сотрудник.
[5] Доктор технических наук, профессор, заведующий лабораторией.

Проблема повышения уровня автономности, живучести, эффективности функционирования сложных технических объектов (СТО), к которым, в частности, относятся перечисленные классы космических средств (КСр), в научной литературе рассматривается в совокупности с решением задач контроля, оценивания и технического диагностирования состояния СТО, проведения реконфигурации (структурной, функциональной, структурно-функциональной реконфигурации) структур СТО, управления ее резервами, альтернативного и многорежимного управления, анализа и синтеза отказоустойчивости и катастрофоустойчивости СТО [2-3]. Однако, к сожалению, все перечисленные исследования носят разрозненный и разобщенный. В связи с этим авторами доклада в выполненном проекте [4] было проведено дальнейшее развитие формируемой ими прикладной теории проактивного управления структурной динамикой СТО, в рамках которой удалось с единых системно-кибернетической позиции подойти к вопросам обеспечения надёжности, живучести, катастрофоустойчивости и отказоустойчивости СТО (в том числе и КСр) сложных объектов на основе конфигурации и реконфигурации их структур [3-4]. Кратко остановимся на основных научных и практических результатах, полученных к настоящему времени при решении различных классов задач реконфигурации СТО, в том числе БА МКА.

Анализ состояния исследований в области решения задач планирования реконфигурации

Анализ показывает, что в общем случае целевое функционирование МКА ДЗЗ может происходить как в расчётных или «нормальных» условиях полёта, так и в аномальных ситуациях, связанных с отказами БС МКА. При этом аномальные полётные ситуации можно условно разделить на два класса: расчётные, или «проектные», последствия которых можно заранее оценить и предусмотреть их парирование; нерасчётные АПС, последствия которых плохо предсказуемы ввиду неопределённости причин, их вызывающих [3-12].

Достижению требуемых значений целевых показателей эффективности (ЦПЭ) МКА препятствуют различные внешние и внутренние возмущающие факторы. Наиболее существенными из внешних факторов, особенно при длительном полёте (период автономного функционирования превышает 5лет) вблизи естественных радиационных поясов Земли (ЕРПЗ), являются воздействия тяжёлых заряженных частиц (ТЗЧ) и высокоэнергетических протонов (ВЭП) внешних излучений (космических, солнечных, ЕРПЗ) с эффектом накопления. Влияние данных факторов ускоряет деградацию комплектующих электро-радиоэлементов для бортовых систем, вызывая их сбои и отказы, изменение текущего технического состояния относительно их исходного (штатного) состояния, что проявляется в виде нештатных или аномальных ситуаций (АС) на борту КА [3-7].

В связи с этим в работах [3-12] подчеркивается, что отказ БС МКА изменяет её структуру, может привести к возникновению и развитию аварийной и кризисной ситуации. Для предотвращения катастрофического поведения системы следует предусматривать меры оперативного реагирования, обеспечивающие безопасность функционирования системы, а также её последующие диагностику и реанимацию, т.е. мероприятия по восстановлению штатных функций. В данном контексте отказоустойчивость (в отличие от безотказности системы, тесно связанной с понятием надёжности) близка, по сути, к понятию живучести системы, поскольку отказ в системе оказывается событием допустимым, а его негативное влияние может быть скомпенсировано посредством реконфигурации соответствующей БС МКА [4-7].

В настоящее время на практике при формировании программ реконфигурации БС МКА используют как уже указывалось ранее, технологию «слепой» реконфигурации, в рамках которой при отказах и нарушениях правильности функционирования соответствующей БС с целью сохранения наиболее приоритетных функций МКА или допустимых условий работоспособности “жертвуют” другими функциями или частью работоспособных элементов. Среди данных вариантов «слепой» реконфигурации [13-23] принято выделять следующие три подхода (варианта) применительно к сложным техническим объектам (СТО), к числу которых относится и МКА.

Вариант I. Функциональные элементы (ФЭ) СТО представляют собой многофункциональные унифицированные (однородные) вычислительные средства. Процедура реконфигурации СТО заключается в следующем [11-14]. Множество решаемых задач разбивается на группы задач с близкими (одинаковыми) характеристиками. Каждая группа задач решается на одном ФЭ.

При отказе ФЭ выполняемая им группа задач передается на ФЭ, где решаются задачи с наиболее низкими приоритетами. Если не удается решить задачи объединенных групп, то задачи группы с низкими приоритетами снимаются с решения.

Вариант II. Более сложная процедура «слепой» реконфигурации представлена в работах [15-19]. В отличие от рассмотренного варианта I ФЭ взаимодействуют посредством некоторой телекоммуникационной подсистемы. Каждый ФЭ содержит процессоры, оперативное и долговременное запоминающее устройство, соответствующие интерфейсы, взаимодействует с внешней средой (объектами управления, операторами и др.) с помощью выделенных для этого аппаратно-программных средств. Возможно наличие в СТО общих ресурсов (внешняя память, базы данных, периферийные устройства и др.).

Пусть данная система осуществляет решение известного множества функций (задач) с заданными требованиями к порядку их выполнения и взаимосвязи согласно выполняемому заданию. При этом каждый из ФЭ способен полностью выполнять одну (специализированный ФЭ) или несколько функций (универсальный ФЭ) из множества функций, возложенных на систему.

Для каждой функции ставится в соответствие некоторое множество полных или упрощенных алгоритмов выполнения функции.

Тогда вариант реконфигурации СТО, обобщающий предложенные подходы в работах [13-17], можно представить следующим образом.

Для начального состояния осуществляется построение плана распределения задач и информационных потоков с учетом технологических, технических, стоимостных, временных, ресурсных и т.п. ограничений. Закрепленные этим планом за ФЭ задачи будем согласно [13-17] называть собственными. В случае перехода СТО (вызванного отказом некоторой совокупности ФЭ) в другое состояние рассматриваются две стратегии.

*Стратегия 1 (*коррекция планов) – собственные задачи и связанные с ними информационные потоки отказавших ФЭ перераспределяются между работоспособными ФЭ, с сохранением за ними выполнения своих собственных задач.

Стратегия 2 (перепланирование) - в новом состоянии осуществляется перераспределение всех решаемых СТО задач.

Обеспечение свойства отказоустойчивости СТО осуществляется путем итогового размещения в различных ФЭ резервных копий алгоритмов решаемых задач. При переходе СТО в другое состояние режим ее функционирования должен быть изменен в соответствии с планом перераспределения задач и информационных потоков для этого состояния, что предполагает активизацию резервных копий алгоритмов решаемых задач работоспособными ФЭ.

Вариант III. При решении задачи реконфигурации функционирования СТО в рамках предлагаемых выше вариантов I и II требуется сформировать множество промежуточных состояний, в которых может находиться СТО. В некоторых случаях это множество может включать, например, все состояния СТО, переход в которые приводит к потере управления наиболее ответственными объектами, либо все состояния, не допускающие перерыва в функционировании СТО из-за необходимости ее перестройки при отказах ФЭ, либо, как правило, все состояния, вероятности перехода в которые из начального состояния не менее заданной (в простейшем случае – все состояния). Учитывая сказанное, возможно и целесообразно задачу реконфигурации СТО в условиях действия случайных возмущений представлять с использованием подходов двухэтапного или многоэтапного стохастического программирования [18-25]. Другими словами, построение начальных планов распределения и планов перераспределения задач и потоков информации осуществляется совместно, т.е. осуществляется выбор начальных планов распределения таким образом, чтобы последующие перераспределения позволили рационально использовать ресурсы СТО, необходимые для компенсации неблагоприятных воздействий [18-21,24].

Следует отметить, что для повышения гибкости (адаптивности) рассматриваемых СТО заблаговременно вносится избыточность во все виды их структур, которая позволяет на практике ставить и решать задачи управления данными структурами (в том числе проводить их реконфигурацию). Однако в ходе «слепой» реконфигурации, как правило, не проводятся следующие операции: учёт и анализ текущих характеристик, решаемых СТО задач и выполняемых функций; анализ и оценивание текущего состояния СТО в целом; оперативный расчёт, оценивание и анализ целевых и информационно-технических возможностей СТО для обоснованного пере-

распределения функций СТО между её работоспособными элементами и подсистемами. Применительно к космической сфере в последние годы появился целый ряд исследований проблем управления восстановлением работоспособности БС МКА, базирующийся на байпасном подходе, связанным с формированием при аварийной полетной ситуации (АПС) обходных путей выполнения основных функций МКА. В указанных ситуациях основная управленческая задача состоит в том, чтобы отыскать приемлемые бортовые ресурсы (вспомогательные, дополнительные контуры управления) и продуктивные способы их использования в АПС для успешного выполнения поставленных перед МКА целевых задач, учитывая различие методов и средств построения их основных контуров [4-10]. При этом живучесть МКА ДЗЗ может обеспечиваться за счёт байпасов, которые формируются не только на основе типовых ресурсов БС (структурных, функциональных, информационных, временных). Наряду с ними могут использоваться ресурсы иного плана, например, естественные ресурсы, т.е. присущие некоторым системам по определению, а также синергетического типа, возникающие при целенаправленном взаимодействии различных систем. В общем случае байпасы могут формироваться на любом уровне иерархической структуры выполнения задач назначения МКА. Например, байпасные каналы в БС, байпасные контуры в составе МКА, резервные МКА как байпасы для выполнения целевых задач космической системы наблюдения. Сущность применения байпасов состоит в том, чтобы с их помощью оперативно осуществлять мероприятия по противодействию проявившимся АПС, их парированию, т.е. нейтрализации последствий, вызванных отказами БС, обеспечивая тем самым живучесть КА ДЗЗ с возможностью реанимации целевого функционирования [4,5,8-10].

Байпасы разрабатываются либо заранее на этапе проектирования штатных средств БС, когда закладываются структурные или функциональные резервы, либо в оперативном порядке после появления АПС. В любом случае их структура и совокупность формируются исходя из концепции многоуровневой байпасности. Прежде всего, это уровни бортовых систем и их компонентов, далее следует уровень МКА как совокупность всех БС, включая бортовой комплекс управления (БКУ). Затем главный уровень – наземный комплекс управления (НКУ), обладающий мощными техническими и управленческими ресурсами, а также возможностями привлечения к решению бортовых задач внешних факторов: искусственных или естественных физических полей; наземных испытательных средств и отладочных комплексов.

Завершая анализ существующих подходов к решению проблем управления реконфигурации БС МКА необходимо отметить, что все они носят достаточно разрозненный разобщенный характер и не позволяют с единых позиций подойти к комплексной автоматизации рассматриваемого управленческого процесса. Вместе с тем, базируясь на разрабатываемой авторами данного доклада прикладной теории проактивного управления реконфигурацией БС МКА, можно конструктивно подойти к решению всех основных задач управления восстановлением работоспособности БС МКА в аварийных и кризисных ситуациях (в том числе и за счет реконфигурации). Рассмотрим конкретное модельно-алгоритмическое и программно-информационное обеспечение, позволяющее решить указанные задачи.

Полученные результаты

При разработке общей методологии и технологий решения задачи реконфигурации БА МКА авторы доклада базировались на двух основополагающих концепциях проведения современных системно-кибернетических исследований. В качестве первой концепции была выделена *концепция комплексного (системного) моделирования* БА МКА, которая предполагает разработку и реализацию новых принципов, подходов к проведению полимодельного описания рассматриваемых КСр, а также разработку и комбинированное использование методов, алгоритмов и методик многокритериального анализа, синтеза и выбора наиболее предпочтительных управленческих решений (в том числе и связанных с их реконфигурацией), связанных с созданием, использованием и развитием рассматриваемых объектов в различных условиях динамически изменяющейся внешней и внутренней обстановок. Для реализации данной концепции при разработке модельно-алгоритмического обеспечения каждого экспериментального образца программного модуля были предложены оригинальные полимодельные описания, а также комбинированные методы и алгоритмы манипулирования с данными описаниями при решении

конкретных прикладных задач расчета, оценивания и анализа показателей надежности и живучести БА МКА в штатных и заданных условиях [3,20-26].

В качестве второй основополагающей концепции была выбрана *концепция проактивного управления* СТО, к числу которых относятся, в том числе, МКА и наземный комплекс управления (НКУ) (последний используется для расширения возможностей по управлению БА МКА в условиях возникновения аварийных и нештатных ситуаций). Проактивное управление, в отличие от традиционно используемого на практике реактивного управления СТО, ориентированного на оперативное реагирование и последующее недопущение инцидентов, предполагает, как уже указывалось ранее, предотвращение возникновения инцидентов за счет создания в соответствующей системе мониторинга и управления принципиально новых прогнозирующих и упреждающих возможностей при формировании и реализации управляющих воздействий, базирующихся на методологии и технологиях системного (комплексного) моделирования [25-27].

В ходе проведенных исследований были предложены следующие основные фазы и этапы решения задачи выбора оптимальных программ управления внутренней и внешней реконфигурации БА МКА (в общем случае активным подвижным объектом (АПО)), которые решались в рамках более общей задачи управления структурной динамикой (УСД) соответствующих КСр. В соответствии с разработанной обобщённой процедурой решения данной задачи на первой фазе должно осуществляться формирование (генерирование) допустимых вариантов многоструктурных макросостояний КСр или, говоря другими словами, должен проводиться структурно-функциональный синтез нового облика КСр, соответствующего складывающейся (прогнозируемой) обстановке.

На второй фазе исследования задачи выбора оптимальных программ реконфигурации (в более широкой трактовке управления структурной динамикой (УСД)) КСр приходится решать целую совокупность частных взаимосвязанных задач многоуровневой и многоэтапной оптимизации, которые представлены на рисунке 1. Кроме того, на данном рисунке приведены разработанные в ходе проекта модели, методы УСД КСр, а полученные научные и практические результаты. К числу последних относятся два экспериментальных образца программных модулей (ЭО ПМ) «Структурная реконфигурация», «Функциональная реконфигурация», в рамках которых решаются соответственно задачи внутренней и внешней реконфигурации. На рисунке 2 изображены указанные два ЭО ПМ в рамках обобщенной структурно-функциональной схемы созданного в ходе проекта экспериментального образца программного комплекса (ЭО ПК), в состав которого вошли и другие программные модули [3].



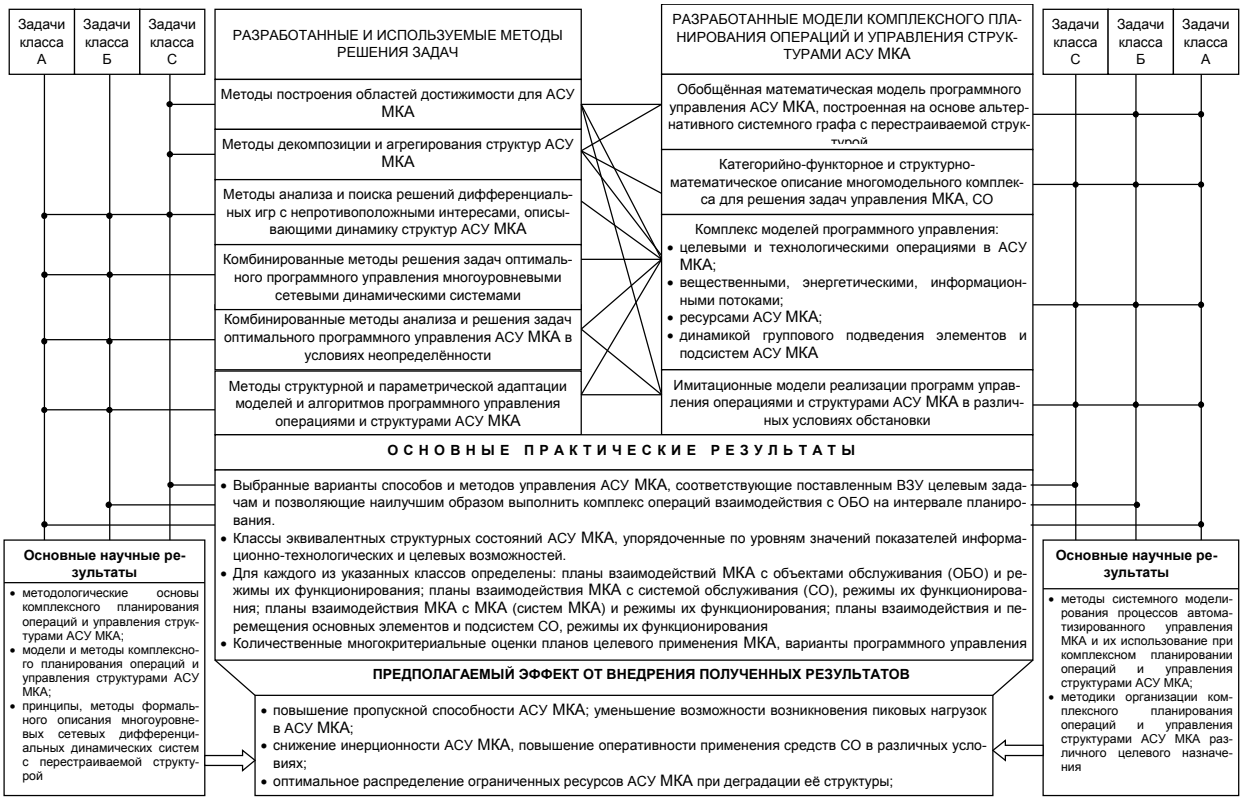


Рис. 1. Основные положения теории управления структурной динамикой АСУ МКА

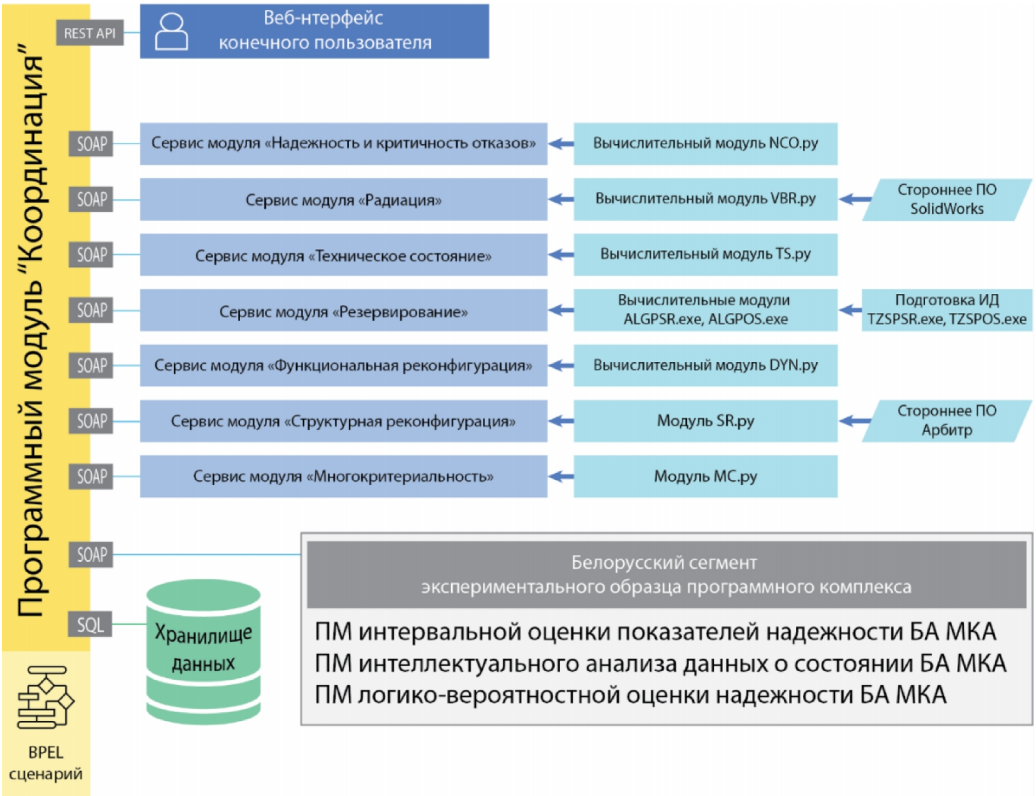


Рис. 2. Структурно-функциональная схема созданного экспериментального образца программного комплекса

В целом, ЭО ПК представляет собой распределённый гетерогенный программный комплекс, основанный на сервис-ориентированной архитектуре. Каждый расчётный модуль разработан в виде веб-сервиса, а их взаимодействие осуществляется посредством сервисной шины. Формирование запросов на выполнение расчётов осуществляется посредством графического пользовательского интерфейса. Графический пользовательский интерфейс функционирует по принципу тонкого клиента и реализован в виде веб-приложения для интернет-браузера на языке JavaScript с разметкой HTML5 и таблицей стилей CSS3. Сервисная шина построена на базе свободно распространяемого программного обеспечения «OpenESB V3.0.5 Standalone Edition». В постоянном режиме функционирует композитное приложение MonitoringSG, обеспечивающее непрерывное функционирование REST интерфейса. В свою очередь, композитное приложение содержит приложение на языке BPEL, формирующее логику работы всего вычислительного процесса. Данное приложение начинает работу при получении запроса от композитного приложения. Сервисная шина исполняется на Linux-сервере с операционной системой Ubuntu Server 17.04. Веб-сервисы программных модулей выполняют запуск и контроль исполнения соответствующих ПМ. В свою очередь, бесперебойная работа самих веб-сервисов обеспечивается системным свободно распространяемым программным обеспечением Circus. Для хранения списка доступных массивов согласованных исходных данных, результатов расчётов и вспомогательной информации используется свободно распространяемая СУБД PostgreSQL 9.6.5. Доступ к таблицам осуществляется также посредством сервисной шины с использованием запросов на языке SQL.

Заключение

Предлагаемое в выполненном проекте рассмотрение вопросов управления реконфигурацией структур БА МКА в общем контексте управления её структурной динамикой позволяет, во-первых, непосредственно связать те общие цели, на достижение которых ориентировано функционирование МКА, с теми целями, которые реализуются в ходе управления структурами МКА и НКУ, во-вторых, обоснованно определить и выбрать соответствующие последовательности решаемых задач и выполняемых операций (действий), связанных со структурной динамикой (другими словами, синтезировать технологию управления МКА), и, в-третьих, осознанно находить компромиссные решения при распределении ограниченных ресурсов, выделяемых на управление структурной динамикой. При этом реализация ЭО ПМ по сравнению с существующими средствами автоматизации поиска вариантов гибкого распределения функций управления между БКУ и НКУ МКА, дают выигрыш в среднем на 15%-20% по показателю полноты реализации ТЦУ МКА и 10%-15% по робастности (устойчивости) сформированных планов реконфигурации.

Исследования, выполненные по данной тематике, проводились при частичной финансовой поддержке грантов РФФИ (№№ 18-08-00419, 16-29-09482-офи-м, введение и первый раздел), РНФ (№16-19-00199 – второй раздел и заключение).

ЛИТЕРАТУРА

1. **Соловьев В.А., Лысенко Л.Н., Любинский В.Е.** Управление космическими полетами. Ч. 1. М.: Изд-во МГТУ им. Н. Э. Баумана, 2009. 476 с.
2. **Лохматкин В.В., Куренко В.И.** Прогнозирование производительности съёмки КА ДЗЗ с учётом надёжности бортовых систем. *Известия Самарского научного центра РАН*. 2013. Т. 15. №4(2).
3. Международный проект – СЧ НИР «Мониторинг – СГ-1.4.1-1» - Разработка методического обеспечения и экспериментального программного комплекса для анализа и прогнозирования надежностных характеристик бортовой аппаратуры маломассогабаритных космических аппаратов на различных этапах жизненного цикла» Заказчик: «НИИ КС имени А.А. Максимова» – филиал ФГУП «ГКНПЦ им. М.В. Хруничева
4. **Ахметов Р.Н., Макаров В.П., Соллогуб А.В.** Принципы управления космическими аппаратами мониторинга Земли в аномальных ситуациях. *Информационно-управляющие системы*. 2012. №1. С. 16-22.
5. **Ахметов Р.Н., Макаров В.П., Соллогуб А.В.** Проблемы реинжиниринга автоматических КА в аномальных полётных ситуациях. *Вестник Самарского государственного аэрокосмического университета*. 2014. № 1(43). С. 9-21.

6. **Ахметов Р.Н., Макаров В.П., Соллогуб А.В.** Байпасность как атрибут живучести автоматических космических аппаратов в аномальных полётных ситуациях. *Вестник Самарского государственного аэрокосмического университета*. 2015. № 4(44). С. 9-21.
7. **Кирилин А.Н., Ахметов Р.Н., Соллогуб А.В., Макаров В.П.** Методы обеспечения живучести низкоорбитальных автоматических КА зондирования Земли: математические модели, компьютерные технологии. М.: Машиностроение, 2010. 384 с.
8. **Борисов А.В., Ковальский Н.П., Любченко Ф.Н.** Перспективные направления создания комплексов приема, обработки и распространения космической информации дистанционного зондирования Земли и связанные с этим проблемы. *Космонавтика и ракетостроение*. 2008. № 5. С. 9—15.
9. **Кирилин А.Н., Аншаков Г.П., Ахметов Р.Н., Сторож Д.А.** Космическое аппаратостроение: научно-технические исследования и практические разработки ГНПРКЦ «ЦСКБ-Прогресс». Самара: Волга—Дизайн, 2011. 280 с.
10. **Ахметов Р.Н.** Модели и методы автономного управления живучестью автоматических космических аппаратов дистанционного зондирования Земли. *Вестник Самарского государственного аэрокосмического университета*. 2008. № 2. С. 194—210.
11. **Кравец В.Г.** Автоматизированные системы управления космическими по-летами / В. Г. Кравец. – М.: Машиностроение, 1995. 256 с.
12. **Кравец В.Г.** Основы управления космическими полетами / В. Г. Кравец, В. Е. Любинский. – М.: Машиностроение, 1983. 224 с.
13. **Богатырев В.А.** Децентрализованное динамическое распределение запросов в многомашинных вычислительных системах / В. А. Богатырев // *Электронное моделирование*. 1994. Т. 16. №3. С. 38–43.
14. **Богатырев В.А.** Децентрализованный метод динамического распределения запросов в отказоустойчивых многомашинных вычислительных системах / В. А. Богатырев // *Автоматика и вычислительная техника*. 1993. №3. – С. 73–75
15. **Бородакий Ю.В., Тарасов А.А.** О функциональной устойчивости информационно–вычислительных систем. *Информационное противодействие угрозам терроризма*. 2006. №7. С. 79–93.
16. **Тарасов А.А.** Функциональная реконфигурация отказоустойчивых систем. М.: Логос, 2012.-152 с.
17. **Турута Е.Н.** К вопросу о построении отказоустойчивых распределенных управляющих систем. *Распределенные системы передачи и обработки информации*. М.: Наука, 1985.
18. **Турута Е.Н.** Обеспечение отказоустойчивости многопроцессорных систем пут□м перераспределения задач отказавших модулей. *Системы управления информационных сетей*. М.: Наука, 1983. С. 183–198
19. **Турута Е.Н.** Обеспечение отказоустойчивости управляющих много-микропроцессорных систем путем перераспределения задач отказавших модулей. *Системы управления информационных сетей*. М.: Наука, 1983. С. 187–198
20. **Павлов А.Н.** Структурный анализ катастрофоустойчивой информационной системы. *Труды СПИИРАН*. 2009. Вып. № 8. С. 28–151.
21. **Павлов А.Н.** Обобщенный алгоритм формирования классов структурных состояний информационных систем. *Информационные технологии и вычислительные системы*. 2009. № 4. С. 3–8.
22. **Павлов А.Н.** Комплексное моделирование структурно-функциональной реконфигурации сложных объектов. *Труды СПИИРАН*. 2013. Вып. № 5. С. 143–168.
23. **Павлов А.Н.** Функциональная реконфигурация чувствительных элементов системы управления движением космического аппарата. *Труды СПИИРАН*. 2013. Вып. № 5. С. 169–181.
24. **Калинин В.Н.** Теоретические основы управления активными подвижными объектами. М.: Министерство обороны СССР, 1974. 130 с.
25. **Калинин В.Н., Соколов Б.В.** Многомодельный подход к описанию процессов управления космическими средствами. *Теория и системы управления*. 1995. №1. С.56–61.
26. **Соколов Б.В.** Комплексное планирование операций и управление структурами в АСУ активными подвижными объектами. СПб.: МО, 1992. 232 с.
27. **Соколов Б.В., Юсупов Р.М.** Комплексное моделирование функционирования автоматизированной системы управления навигационными космическими аппаратами. *Проблемы управления и информатики*. 2002. № 5. С. 1–15.

A.U. Kulakov, F.M. Kulakov, B.V. Sokolov, A.N. Pavlov, S.A. Potryasaev (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences)

Methodological, methodical, and technological basis of description and solving satellite-borne equipment small- mass-dimension spacecraft reconfiguration problems is proposed. The complex technical objects structure dynamics applied theory are suggested for investigation and solving reconfiguration problems. Models, methods, algorithms for satellite-borne equipment small- mass-dimension spacecraft reconfiguration are developed. These models, methods, algorithms are based on conception of integrated modelling and proactive control.

ТИПОЛОГИЯ СИСТЕМНЫХ ЭФФЕКТОВ МНОГОФАКТОРНЫХ ВОЗДЕЙСТВИЙ В ОТКРЫТЫХ СИСТЕМАХ

Т. Л. Качанова[1], **Б. Ф. Фомин**[2], **О. Б. Фомин**[3], **В. О. Агеев**[4]
(*СПбГЭТУ «ЛЭТИ», С.-Петербург, ул. Проф. Попова, 5, bfomin@mail.ru*)

Аннотация

Задача о системной (эмерджентной) реакции каждого показателя на множественное воздействие десятков (сотен) влияющих факторов в открытой системе решена на основе научно-достоверного знания об общей онтологии открытой системы, автоматически генерируемого технологиями физики открытых систем непосредственно из описания системы большим многомерным массивом (сотни, тысячи показателей) гетерогенных эмпирических данных. Метод решения определяет типологию изменчивости каждого реагирующего показателя в ответ на любое многофакторное воздействие.

Ключевые слова: физика открытых систем, наука о данных, онтология открытых систем, знание-центрическая системная аналитика, эффекты многофакторных воздействий

Введение

Междисциплинарное научное направление *физика открытых систем* (далее – ФОС) возникло на базе статистической физики и синергетической парадигмы [1-3]. ФОС исследует динамические и статистические описания сложных движений и релевантные методы их моделирования [4-6]. Творческие активы ФОС – физика и математика. Системы в ФОС – математические динамические модели непрерывного или дискретного времен, открытость систем – принцип, сложность систем – сложность движения. Предмет ФОС – общие идеи теории сложности, сверхзадача – научное понимание сущности сложности и рациональное объяснение глубокой взаимосвязи сложности с законами природы.

В середине 90-х возникло *новое направление ФОС*, в рамках которого идет становление *киберфизической парадигмы системологии многомерной системной знание-центрической аналитики* [7, 8]. Ее цель – глубокое исследование природных, общественных, антропогенных и сложных технических открытых систем, исходно заданных большими массивами многомерных гетерогенных эмпирических данных [9, 10]. На этом направлении ФОС преодолела фундаментальный барьер сложности открытых систем и создала собственный научный аппарат. Ее подходы и методы воплотились в информационные и когнитивные технологии, обеспечившие:

- автоматическое производство научно-достоверного онтологического знания об открытых системах с сотнями и тысячами показателей из больших массивов гетерогенных эмпирических данных без обращения к экспертному знанию, субъективному анализу и интерпретациям [11, 12];
- научное понимание и рациональное объяснение общей онтологии открытых систем [13];
- анализ ценности (правильности, полноты, завершенности) знания об онтологии систем [14-16].

В докладе используются достижения *только этого направления ФОС*. Становление аналитики ФОС идет по пути решения *общих системных задач на основе онтологического знания* [8, 17]. В качестве таких задач мы рассматриваем: дифференциальную экспрессию генов по данным, полученным от микрочипов; естественное классифицирование; типологию системных эффектов многофакторных воздействий; идентификацию состояний, событий и ситуаций; прогноз состояний, событий, ситуаций и изменений; системную компаративистику. Для трех первых задач из этого списка методы и технологии автоматического решения разработаны до полной готовности. Уровень готовности трех других задач – опытный образец, работающий с реальными данными. Методы решения задач на основе знания общей онтологии открытых систем создаются по единому сценарию. В этом сценарии каждая общая системная задача проявляет

[1] Доктор технических наук, профессор
[2] Доктор технических наук, профессор
[3] Ассистент
[4] Ассистент

себя по-своему. Сложность ее решения каждый раз преодолевается ФОС на основе знания общей онтологии систем.

Автоматическое производство, хранение, эксплуатацию и обращение системного знания выполняет Технологическая платформа ФОС (далее – Платформа). В ее составе: аналитическое ядро (далее – АЯ ФОС), дескриптивный, конструктивный и проективный компоненты [8, 17, 18]. *АЯ ФОС – интеллектуальная «машина»* Платформы. Она автоматически производит научно-достоверное онтологическое знание о системе из ее эмпирического описания, автоматически исследует правильность, полноту, завершенность полученного знания, автоматически создает на основе знания информационный, интеллектуальный, когнитивный и технологический ресурсы (далее – ИИКТ-ресурсы) *системного знания*. *Дескриптивный компонент* формирует многоаспектное многоцелевое видение системы, создает ее полный эмпирический контекст, в соответствие с которым «сырые» исходные данные о системе и системных проблемах преобразуются технологиями АЯ ФОС в нормативно организованную эмпирическую систему данных, отвечающую целям и задачам исследования. *Конструктивный компонент* переводит системное знание, полученное АЯ ФОС, в предметный формат и создает *ИИКТ-ресурсы решений* общих системных задач с учетом их проявления в исследуемых предметных областях. *Проективный компонент* использует рабочие продукты технологий конструктивного компонента для построения на их основе *RTD-кластеров* многомерной знание-центрической аналитики открытых систем *в предметных доменах*: «Безопасность» (химическая, радиационная, социальная); «Системная биология»; «Медицина, здравоохранение»; «Планетарная и солнечно-земная физика»; «Системная инженерия». RTD-кластеры – интеллектуальные сервисы для автоматизации R&D и автоматической подготовки знание-центрических решений конкретно-предметных системных проблем.

Задача о системных эффектах

Эта задача поставлена и решена нами как *общая задача о типологии изменчивости реагирующих показателей в ответ на многофакторные воздействия в открытых системах* [8, 19]. Ее сложность характеризуют: большая размерность исходных данных (сотни и тысячи показателей); гетерогенность данных; большое количество влияющих факторов и реагирующих величин (десятки, сотни); разнообразие способов шкалирования величин; наличие пропущенных значений данных. Знание о закономерных отношениях между факторами влияния и реагирующими показателями *имплицитно содержатся в знании об общей онтологии системы «Поле многофакторных эффектов»* (далее – «ПМЭ»). Каждый объект «ПМЭ» задает одно конкретное актуальное состояние *системы «ПМЭ»*, представленное вектором значений ее показателей. В этом векторе каждому конкретному заданному набору значений действующих факторов соответствует набор конкретных значений реагирующих показателей. Вектор показателей (единый для всех объектов «ПМЭ») является носителем исходного эмпирического описания системы «ПМЭ». Технологии ФОС выявляют *полное множество собственных качеств системы «ПМЭ»*, преобразуют ее собственные качества в пропозициональные структуры и предикаты *сигнификативной схемы* «ПМЭ», на основе которой определяют *денотативную схему* «ПМЭ». Типология системных эффектов выявляется в результате *многокритериальной дифференциации* этих схем.

Концепция решения

Идеей решения задачи является познание общей онтологии системы «ПМЭ» и получение на ее основе онтологии системных эффектов многофакторного воздействия (Рис. 1). Система «ПМЭ» предопределяет типологию многофакторных эффектов (связь «Импликация»). Связь «Экспликация 1» означает, что собственные качества системы «ПМЭ» преобразуются в пропозициональные структуры и предикаты сигнификативной схемы «ПМЭ», которая, в свою очередь, определяет денотативную схему «ПМЭ». Связь «Экспликация 2» выражает развитие, детализацию, уточнение и конкретизацию денотативной схемы «ПМЭ» путем ее погружения в системный контекст. Сигнификативная и денотативная схемы представляют знание об онтоло-

гии «ПМЭ». *Сигнификативная схема* задает общий вид прототипического воздействия для каждого реагирующего показателя. На его основе определяется спектр предикатов, объясняющих любой конкретный вариант воздействия и эффект на это воздействие. *Денотативная схема* задает модели всех фактических разновидностей эффектов на воздействия и определяет типологию ответов системы на эти воздействия.

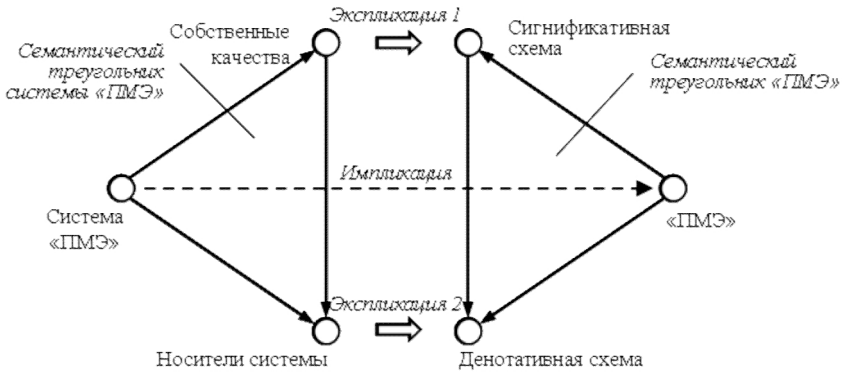


Рис. 1. Схематическое представление концепции решения

Многокритериальная дифференциация эффектов использует комбинацию значений критериев, характеризующую конкретную моделируемую ситуацию и системную значимость атрибутов, оценивающих показатели и эталоны в аспектах *статальности* и *акциональности*, а также полный системный контекст этой ситуации.

Модель решения

Концепция решения преобразуется в *модель метода решения* (далее – модель). Модель определяет и выражает смысл понятия «Системный эффект многофакторного воздействия» в категориях «Представление «ПМЭ»» и «Онтология «ПМЭ»» (Табл. 1).

Таблица 1.

Модель метода решения

Градация категории «Онтология «ПМЭ»»	Градация категории «Представление «ПМЭ»»		
	Целое	Частное	Конкретное
Онтология системы	Системная предикация	Детерминация уровней величин	Системная градация величин
Интенсионал	Семантическая связь	Прототипический сигнификат	Входы-выход
Экстенсионал	Смысловая доминанта эффекта	Актуальный контент эффекта	Объект

Исходными данными модели являются онтология системы, набор действующих величин, набор реагирующих величин. Модель раскрывает структуру семантического, денотативного и оценочного компонентов «ПМЭ». Осмысление многофакторных системных эффектов осуществляется от градации «Конкретное» к градации «Целое» и от градации «Онтология системы» к градации «Экстенсионал».

Градации категории «Представление «ПМЭ»» задают разные *формы выражения смыслов системных эффектов* многофакторного воздействия. Градация «Конкретное» характеризует показатели состояния системы и ее окружения в каждом актуальном состоянии через диспозицию «Входы – выход». Градация «Частное» выявляет неделимые смысловые единицы онтологии «ПМЭ». Таковыми являются: детерминированные уровни значений величин; наборы уровней значений действующих факторов; значения реагирующих показателей, характерные для уровней значений; актуальные варианты системных эффектов. Градация «Целое» характеризует онтологию «ПМЭ» в целом через полные множества: эталонов состояний собственных качеств системы «ПМЭ»; смысловых связей между действующими факторами и реагирующими показателями; типовых системных эффектов многофакторных воздействий.

Градации категории «Онтология «ПМЭ»» означают этапы *смыслового анализа системных эффектов*. Градация «Онтология системы» задает базу для формирования онтологии системных эффектов. Градация «Интенсионал» раскрывает устроение сигнификативной схемы эффектов воздействий. Градация «Экстенсионал» характеризует денотативную схему эффектов.

Понятия модели:

- *Системная градация величин* – полный набор примитивов, предназначенных для описания значений показателей состояния системы и среды в единой шкале измерений; это выявляет оппозицию большого и малого уровней значений величин, установленную самой системой.
- *Детерминация уровней величин* – полные множества системных моделей, статистически различающих уровни значений конкретных показателей. Они задают статистические связи между показателями и эталонами собственных качеств системы «ПМЭ», которые объясняют актуальные состояния системы.
- *Системная предикация* – полное множество статистически выявленных отношений «Эталон состояний - уровень значения величины». Различительная способность эталонов может быть слабой (обусловлена статистически) или сильной (обусловлена системой и статистически).
- *Входы - выход* - диспозиция действующих факторов и реагирующего показателя. Набор действующих факторов задан сообразно целям исследования. Всякий реагирующий показатель выбран из целевого множества показателей, типология изменчивости которых должна быть раскрыта.
- *Прототипический сигнификат* - набор действующих показателей, чье влияние значимо и которые принадлежат множеству входов, детерминирующих изменчивость уровней значений каждого реагирующего показателя.
- *Семантическая связь* - конкретный вид отношений между действующим и реагирующим показателями. Эти отношения раскрывают виды форм воплощения прототипического сигнификата, который задает инвариантное отношение между уровнями действующих величин и уровнем реагирующей величины.
- *Объект* – конкретный объект пространства решений.
- *Актуальный контент <многофакторного> эффекта* – конкретная форма носителя закономерности, объясняющей изменчивость реагирующего показателя.
- *Смысловая доминанта <многофакторного> эффекта* – семантический фильтр, основанный на оппозициях признаков актуальных контентов, структурирующий денотативную область многофакторных эффектов.

Метод решения

Типологию эффектов для любого и каждого реагирующего показателя при заданном наборе действующих показателей определяет метод решения задачи (далее – метод). Каждый тип эффекта рационально объясняют характерные наборы эталонов состояний собственных качеств системы. Для каждого реагирующего показателя установлена типология его системного ответа. Она задана набором значимых факторов влияния, которые получены из идеального образца отношения «Многофакторное воздействие – системный эффект».

Понятиям модели отвечают вычислимые объекты метода (Таблица 2):

Таблица 2

Связь понятий модели с вычислимыми объектами метода

Понятие модели	Объект метода
Системная градация величин	Система в градуированных данных
Детерминация уровней величин	Показатель в системном целом
Системная предикация	Актантная структура эталона
Входы – выход	Целевой список показателей
Прототипический сигнификат ()	Идеальный образец отношения «Воздействие – эффект»
Семантическая связь	Ролевая семантика предиката
Объект	Носитель эффекта
Актуальный контент многофакторного эффекта	Модель системного эффекта
Смысловая доминанта многофакторного эффекта	Макроструктура эффекта

Система в градуированных данных. В онтологическом знании разработана специальная качественная 17-пунктная шкала значений показателей, обеспечивающая моделирование изменяющихся величин. Эта шкала редуцируется к 5-, 3-, 2-пунктным шкалам уровней значений величин. При определении формального объекта «Система в градуированных данных» используется 3-пунктная шкала («Высокий уровень», «Средний уровень», «Низкий уровень»). В результате такого шкалирования форма представления системы ее эмпирическим описанием в виде таблицы «Объект-свойство» заменяется новой формой, в которой наблюдаемые значения показателей преобразованы в значения на 3-пунктной шкале.

Показатель в системном целом. В онтологическом знании уровень значений каждого показателя определен всем системным целым в рамках реконструкции каждого конкретного актуального состояния. Каждая реконструкция характеризуется уникальным набором эталонов, проявляющих внутрисистемные механизмы, детерминирующие уровни значений показателей, и имеет рациональное объяснение. Уровень значения показателя устанавливается и различается системой в целом через типические для этого уровня эталоны. Каждый эталон как состояние собственного качества системы является частью системного целого и всем целым в условиях этой части.

Актантная структура эталона. Эталон является атрибутивным словом, если он выявляет набор атрибутов (конкретных уровней значений показателей) всего системного целого. Число мест в наборе атрибутов (валентность) – активность эталона. Активность определяется посредством статистических обобщений. Если при этом показатель из набора атрибутов входит как элемент в состав эталона, то он является системно обусловленным атрибутом.

Целевой список показателей. В списке представлены имена показателей, рассматриваемых в качестве действующих (*задается экспертами*). В каждой реконструкции действующие показатели заданы конкретным набором уровней их значений. Комбинация уровней значений действующих показателей в отдельно взятой реконструкции – один вариант многофакторного воздействия. Множество таких вариантов на всем множестве реконструкций – *поле многофакторных воздействий*. В каждой реконструкции реагирующий показатель (*задается экспертами*) имеет конкретный уровень значения (реакция на вариант многофакторного воздействия). Все множество уровней значений реагирующего показателя – *поле системных эффектов на многофакторные воздействия*.

Идеальный образец отношения «Воздействия – эффект». Для реагирующего показателя устанавливается множество эталонов, актантные структуры которых содержат этот показатель. В каждой из этих актантных структур выявляется наличие таких действующих показателей, для которых эталон, соответствующий данной структуре, становится реляционным словом языка систем (*эталоном-предикатом*). В этой роли эталон-предикат устанавливает отношение между конкретным действующим показателем и реагирующим показателем. Это отношение имеет статистическое подтверждение. В итоге формируется список значимых действующих показателей, ответственных за системные эффекты реагирующего показателя. Для каждого действующего показателя определяется знак статистической связи с реагирующим показателем. Каждый уровень значений реагирующего показателя рассматривается как прототип системного эффекта на воздействия и является *прототипическим денотатом эффекта*. Для этого денотата вводится *прототипический сигнификат*, представляющий собой типический набор уровней значений действующих показателей.

Ролевая семантика предиката. Семантическое описание эталонов-предикатов предполагает их классификацию по ролевой семантике. Класс предиката определяют ролевые функции его актантов и смыслообразующая функция самого предиката в системном целом. Ролевая функция актанта, представляющего реагирующий показатель, заключается в носительстве конкретного уровня его величины. Ролевая функция актанта, представляющего действующий показатель, принимает два значения – отвечает (не отвечает) идеальному образцу прототипического сигнификата. Для предиката, являющегося системно обусловленным относительно реагирующего показателя, имеются два значения его смыслообразующей (ролевой) функции: эталон представляет *системный механизм изменчивости (акциональности)*; эталон представляет *системный механизм статальности*. Значения всех ролевых функций актантов и предиката задают пять классов эталонов-предикатов:

- класс «Действие» характеризуется проявлением системного механизма акциональности реагирующего показателя. Во всех других классах эталон-предикат проявляет системный механизм статальности;
- класс «Состояние» определяется соответствием уровня действующего показателя идеальному образцу;
- класс «Отношение» устанавливает системно-закономерную связь между действующим показателем и реагирующим показателем;
- класс «Признак» отвечает условию несоответствия уровня действующего показателя идеальному образцу (объясняет разнообразие воздействий при неизменном эффекте);
- класс «Свойство» близок классу «Состояние», однако в нем не детерминируется уровень значения реагирующего показателя.

Носитель эффекта. Этот объект представлен совокупностью фрагментов реконструкций актуальных состояний системы, охватывающих только действующие показатели и реагирующий показатель с их уровнями значений, а также набор эталонов, детерминирующих эти уровни.

Модель системного эффекта. Данный объект представляет собой фреймовую структуру – стереотипное описание события «Воздействие-эффект». *План выражения модели* актуализирует конкретный системный эффект на определенный вариант воздействия. *План содержания модели* раскрывает внутрисистемные механизмы, объясняющие возникновение и характер этого конкретного эффекта. Модель характеризует и объясняет конкретный системный эффект на определенный вариант воздействия *как уникальный научно установленный факт* в полном системном многообразии возможных эффектов. Актуализация контента осуществляется на каждом носителе конкретного эффекта путем указания:

- варианта воздействия через уровни значений действующих показателей;
- уровня значения реагирующего показателя;
- идеального образца воздействия;
- класса ролевой семантики для каждого эталона-предиката, отражающего определенное семантическое отношение между предикатом и его окружением, а также значимости этой ролевой семантики в аспекте «Динамика/статика»;
- атрибута направленности/ненаправленности действия для каждого эталона-предиката;
- главной ролевой функции (в аспекте «Динамика/статика») и ее значимости для каждого действующего показателя;
- атрибута подвижности/неподвижности для реагирующего показателя.

Макроструктура эффекта. Этот объект метода выявляет закономерное единство во всем актуальном системном многообразии эффектов воздействий и задает их типологию. Типология определяется путем назначения и разрешения набора бинарных семантических оппозиций. Каждая оппозиция предъявляет различие между элементами *плана выражения события в модели системного эффект*а. Этому различию соответствует различие между элементами *плана содержания события в модели эффекта*. Конструктивными элементами для оппозиционного анализа в рамках модели системного эффекта являются: вариант воздействия, потенциал подвижности реакции, классы ролевой семантики предикатов, эталоны с системно обусловленным атрибутом. Различие в проявлениях этих конструктивных элементов оценивается через оппозиции:

- «*Статальность/акциональность*»: разрешается путем присвоения каждому предикату модели эффекта атрибута статальности или акциональности с указанием его статистической значимости для системных эффектов и интеграцией полученных атрибутов по всем предикатам модели в общую оценку, характеризующую степень стабильности системной реакции на воздействие; играет ведущую роль в определении типа эффекта; является эквиполентной оппозицией (члены оппозиции равноправны, каждый член имеет собственный признак); для каждого дифференциального признака «Статальность» и «Акциональность» независимо вводится градуальная оппозиция «Сильная/слабая» (степень выраженности признака).
- «*Релевантность/нерелевантность*»: разрешается путем сопоставления варианта многофакторного воздействия с прототипическим образцом и нахождения меры их соответствия, характеризующей вариабельность действующих показателей;

− «*Обусловленность/необусловленность*»: разрешается посредством выявления в имеющемся наборе эталонов-предикатов множества системно обусловленных уровней всех действующих показателей, характеризующих полноту модели эффектов; привативная оппозиция (один из ее членов обладает дифференциальным признаком, другой не обладает).

− «*Ненаправленность/направленность*»: разрешающуюся путем анализа выявленных системных механизмов действия, характеризующих способность системного целого изменить уровень реагирующего показателя; привативная оппозиция.

В результате *оппозиционного анализа* возникает *пучок дифференциальных признаков с иерархической организацией*. На его основе выявляется типология системных многофакторных эффектов. Каждый тип эффекта конструируется на основе характерной композиции значений дифференциальных признаков. Полученная типология эффектов определяет семантическую макроструктуру денотативной схемы. Денотативная область не является прямым отражением конкретики системных эффектов, а есть когнитивное отражение установленных типов системных эффектов.

Функциональное описание метода

Отношения между формальными объектами метода наследуют отношения между понятиями модели. Конструктивные формы отношений порождения и преобразования объектов определяют функциональность метода (Рис. 2).



Рис.2. Функциональная схема метода:
1 – производство системного знания; 2 – экспертное назначение; 3 – шкалирование; 4 – локализация эффекта; 5 – дистрибуция показателей; 6 – определение валентностей; 7 и 8 – прототипирование; 9 – ролевое оценивание; 10 и 11 – заполнение фрейма; 12 – выявление типологии эффектов

Производство системного знания об онтологии системы «ПМЭ»: обеспечено технологиями АЯ ФОС. *Назначение целевых показателей*: эксперты указали действующие факторы и назначили реагирующий показатель. *Представление системы в градуированных данных*: шкалы измерения уровней значений переменных преобразованы в шкалы, градуированные самой системой. *Локализация эффекта*: получены модели реконструкций всех актуальных состояний системы (в них представлены все системные эффекты многофакторных воздействий). *Дистрибуция показателей*: на основе моделей реконструкций получены частоты сочетаний показателей и эталонов состояний системы, определена статистическая значимость этих сочетаний. *Определение валентностей*: семантически детерминировано доминирование эталонов состояний; установлена структура отношений эталонов и актантов; выявлены валентности актантов. *Прототипирование:* установлен список значимых действующих факторов, вызвавших системный эффект реагирующего показателя; для каждого действующего показателя определен знак ста-

тистической связи с реагирующим показателем; найдены *прототипические денотаты и сигнификаты* эффекта.

Ролевое оценивание: для всего множества референтов отношения «Воздействие – системный эффект» установлен состав *классов ролевой семантики* эталонов-предикатов (каждый класс выделяет определенный *аспект ситуации* системной реакции на воздействие). *Заполнение фрейма*: для каждой конкретной ситуации «Воздействие – системный эффект» вычислен набор атрибутов, характеризующих моделируемую ситуацию в аспектах *статальности и акциональности. Выявление типологии эффектов*: проведены многокритериальная дифференциация и типологический анализ системных эффектов.

Пример апробации метода

Рестеноз и тромбоз стента после чрезкожного коронарного вмешательства (ЧКВ) на коронарных артериях является важнейшей клинической и социально-экономической проблемой высокотехнологичного лечения ишемической болезни сердца (ИБС). Особенности развития рестеноза и тромбоза внутри стента и формирование индивидуальной чувствительности к действию антитромбоцитарных препаратов в каждом конкретном случае определяется сочетанием большого числа различных патогенетических факторов (клинических, анатомо-морфологических, молекулярно-генетических, биохимических, технических) [20]. При апробации метода нами использован многомерный массив данных, полученных в результате обследований больных ИБС с острым коронарным синдромом после ЧКВ со стентированием в условиях стационара «Федерального центра сердца, крови и эндокринологии им. В. А. Алмазова» (Санкт-Петербург). Решение задачи интервенционной кардиологии опубликовано нами в статье [21]. Полученные результаты показали преимущества метода: охват исследуемой проблемы в ее естественном масштабе и реальной сложности; учет большого количества самых разнообразных факторов, влияющих на развитие рестеноза; создание целостной картины патофизиологии рестеноза и тромбоза интракоронарных стентов у больных ИБС.

Заключение

Для любого реагирующего показателя метод гарантирует: получение прототипического сигнификата; построение прототипического денотата; определение полного набора типов системных эффектов на любые возможные варианты многофакторных воздействий. Метод позволяет получить для каждого референта «ПМЭ» новый формат представления. В этом формате фиксируется вариант воздействия и набор типов системных ответов, установленных соответствующей моделью и макроструктурой эффекта.

На полном множестве референтов существует проблема типологии глобальной системы. Ее решение даст ответ о системной реакции всего множества реагирующих показателей на любое многофакторное воздействие. Это – сверхзадача качественной теории открытых систем. Изложенный метод является важным шагом к ее успешному решению.

ЛИТЕРАТУРА

1. **Haken H.** Synergetics, an Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology. New York: Springer-Verlag, 1983. 336 p.
2. **Климонтович Ю. Л**. Введение в физику открытых систем. М.: Janus-K, 2002. С. 284.
3. **Климонтович Ю. Л.** Статистическая теория открытых систем. Т. 1. М.: Janus, 1995. С.608
4. **Чернавский Д. С.** Синергетика и информация. Динамическая теория информации. М.: Наука, 2001. 244 с.
5. **Малинецкий Г. Г.** Хаос. Структуры. Вычислительный эксперимент. Введение в нелинейную механику. М.: Эдиториал УРСС, 2000. 256 с.
6. **Пригожин И.** Конец определенности: Время, хаос и новые законы природы. Ижевск: НИЦ «Регулярная и хаотическая динамика». 2000. 208 с.
7. **Kachanova T., Fomin B.** Physics of Systems is a postcybernetic paradigm of systemology // Intern. Symp. “Science 2.0 and Expansion of Science: S2ES” in the context of the 14th World-Multi-Conference “Systemics, Cybernetics and Informatics”: WMSCI 2010. Orlando, USA, 2010. C. 244-249.
8. **Kachanova T., Fomin B., Fomin O.** Generating scientifically proven knowledge about ontology of open systems. Multidimensional knowledge-centric system analytics. *Ontology in Information Science*.—InTechOpen, 2018. P. 169–204.

9. **Качанова Т. Л., Фомин Б. Ф.** Основания системологии феноменального. — СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 1999. 180 с
10. **Качанова Т. Л., Фомин Б. Ф**. Метатехнология системных реконструкций. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2002. 360 с.
11. **Качанова Т. Л., Фомин Б. Ф.** Технология системных реконструкций. СПб.: «Политехника», 2002. 146 с.
12. **Kachanova T., Fomin B.** Cognition of ontology of Open Systems. *Procedia Computer Science Journal*.— 2017. P. 339—346.
13. **Качанова Т. Л., Фомин Б. Ф.** Введение в язык систем. СПб.: «Наука», 2009. 340 с.
14. **Качанова Т. Л., Фомин Б. Ф**. Методы и технологии генерации системного знания. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2012. 132 с.
15. **Kachanova T., Fomin B.** Physics of open systems: Generation of system knowledge. *J. Systemics, Cybernetics and Informatics*. 2013. V. 11/2. C. 73-82.
16. Качанова Т. Л., Фомин Б. Ф. Квалитология системного знания. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2014. 132 с.
17. **Качанова Т. Л., Фомин Б. Ф**. Знание-центрическая аналитика многомерных открытых систем/ Материалы 9- й Конференции “Информационные технологии в управлении (ИТУ-2016)”. Октябрь 04-06; Россия. СПб.: Государственный научный центр Российской федерации АО «Концерн «ЦНИИ «Электроприбор». 2016. С. 121-129.
18. **Качанова Т. Л., Фомин Б. Ф. Фомин О. Б.** Аналитическое ядро физики открыты систем. *Известия СПбГЭТУ «ЛЭТИ»*. 2018. Вып. 3. С. 5-12.
19. **Качанова Т. Л., Фомин Б. Ф.** Системные эффекты многофакторных воздействий в открытых системах. *Известия СПбГЭТУ «ЛЭТИ»*. 2017. Вып. 1. С. 28-38.
20. **Березовская Г. А. Ганюков В. И., Карпенко М.А.** Рестеноз и тромбоз внутри стента: патогенетические механизмы развития и прогностические маркеры. *Рос. Кардиологический журная*. 2012; 6(98). С. 91-95.
21. **Качанова Т. Л., Фомин Б. Ф.** Применение метода определения типологии системных эффектов многофакторных воздействий. *Известия СПбГЭТУ «ЛЭТИ»*. 2017. Вып. 2. С. 19-29.

T.L. Kachanova, B.F. Fomin, O.B. Fomin, V.O. Ageev (Saint-Petersburg Electrotechnical University SPbETU “LETI”, Saint-Petersburg)

Typology of system effects arising from multifactorial influences in open systems

The problem about system (emergent) response of each variable in open system to multiple influence of tens (hundreds) of acting factors is solved on the basis of scientifically proven knowledge about general ontology of open system. This knowledge is automatically produced by technologies of physics of open systems, directly from description of the system that is represented by a big multidimensional set (hundreds, thousands variables) of heterogeneous empirical data. The solution method defines typology of variability of each reacting variable as a response to any multifactorial influence.

УПРАВЛЕНИЕ КОНФРОНТАЦИОННЫМ ВЗАИМОДЕЙСТВИЕМ В УСЛОВИЯХ РЕСУРСООБМЕНА СТОРОН - ДИНАМИЧЕСКАЯ МОДЕЛЬ

В. М. Шишкин[1],
(*СПИИ РАН, С.-Петербург, 14-я линия В.О., д. 39, vms@iias.spb.su*)

К. Е. Колесников[2]
(*Главное управление Минюста России по Санкт-Петербургу, Исаакиевская пл., д. 11, kolesnikovib@gmail.com*)

Аннотация

Представлены результаты исследования по одному из направлений развития дифференциальной модели противоборства. Первоначально стороны в ней рассматривались как автономные системы, взаимодействие которых ограничивалось управлением угрозами безопасности в отношении противника. Добавлен ещё один канал взаимодействия через ресурсообмен, делающий стороны не только конфротирующими, но и взаимозависимыми, что изменяет динамику переменных и возможности управления процессом, позволяя, в частности, реализовать самостоятельную алгоритмическую надстройку в виде принуждающего регулятора-гипервизора. Предложены условия его существования, критерии работы и изменения в основных уравнениях модели.

Ключевые слова: дифференциальные уравнения, противоборство, управление, ресурсообмен, гипервизор.

Введение

Изучение процессов конфронтации в разных предметных областях и сферах деятельности представляет интерес исследователей уже давно. Спектр работ на эту тему достаточно широк, начиная от абстрактных математических моделей с использованием различного научно-методического аппарата, до прикладных технических решений, не говоря о многочисленных вербальных исследованиях психологической или политологической направленности. В последнее время значительное внимание уделяется вопросам противоборства в информационной сфере, причём также в широком спектре приложений и интерпретаций.

Здесь хотелось бы отметить некоторые работы, характеризующие разнообразие подходов к данной проблеме. Например, крайне актуальные для военно-технического применения вопросы теории радиоэлектронной борьбы (РЭБ) и так называемых сетецентрических войн в комплексе с психологическими аспектами деятельности операторов [1, 2] или не менее актуальные социотехнические задачи управления и противоборства в социальных сетях [3]. С другой стороны – математико-философские исследования состояния и перспектив информационного взаимодействия с точки зрения безопасности в условиях конфронтации [4, 5].

Надо сказать, что прототип представляемой модели первоначально также был ориентирован на информационную сферу, но в системном взаимодействии с социально-экономическим развитием и обеспечением национальной безопасности - была формализована как система дифференциальных уравнений иллюстративная процессная схема из [6].

Таким образом изначально модель оказалась в некотором смысле интегральной, связывающей во взаимодействии существенно разнородные и трудно соизмеримые факторы, что выделяет её из предметно-ориентированных подходов, делая уязвимой для критики, но одновременно, как показал опыт, допускающей наряду с абстрактными, различные интерпретации, в том числе, предметно-ориентированные, с получением полезных результатов.

Представляемая работа является продолжением серии исследований на более поздней её версии, динамической модели противоборства (ДМП), проведённых за последние два года, некоторые результаты которых приведены в [7]. Несмотря на предельную простоту модели, содержательная интерпретация результатов вычислительных экспериментов, не противоречащая,

[1] Кандидат технических наук, старший научный сотрудник
[2] Специалист

по крайней мере, здравому смыслу, показала, что минимизированный набор элементов в модели и логика отношений между ними были определены достаточно разумно. Модель в конечном итоге была сформулирована в универсальных абстрактных терминах, но, как показано, например, в [8], при необходимости адаптируется к различным предметным областям, в частности, к процессам информационной конфронтации, в том числе, в технических приложениях, и масштабируема, при этом эквивалентные элементы разнородных противоборствующих сторон содержательно могут формулироваться неодинаково.

Элементный состав и структура одного из исходных вариантов [9], прототипа рассматриваемой модели, не предполагали наличие стороны противника и представляли собой автономную закрытую систему с одной, в некотором смысле критериальной, переменной, по которой оценивалось состояние системы в целом – уровнем безопасности. Такой выбор был естественным, так как модель разрабатывалась в контексте анализа проблем безопасности. Однако без учёта среды, внешних воздействий или взаимодействий, эта модель, несмотря на позитивные результаты экспериментов и показанную в эксперименте обнадёживающую правдоподобность поведения, выглядела слишком идеально.

Тогда из разнообразных вариантов учёта и формализация внешних воздействий для определённости и потенциальной прагматичности (наметилась тенденция на конфронтацию в различных предметных областях и масштабах) предпочтение было отдано модели двустороннего конфронтационного взаимодействия. В качестве сторон в модели выступали две структурно эквивалентные подсистемы, подобные исходной, автономной, взаимодействие которых ограничивалось возможностью влиять на уровень угроз у противника с последующим изменением его уровня безопасности в целом.

При этом возникла необходимость управления процессами в подсистемах, поскольку стороны функционировали независимо и должны были с учётом состояния противника следовать своему целеполаганию. В качестве инструмента управления в модель для каждой из сторон был включен блок управления и добавлена ещё одна переменная – ресурс, понимаемый в интегральном смысле, пополняемый за счёт основной, производящей деятельности. Функция блока управления состоит в распределении долей имеющегося ресурса. В качестве целей управления по отношению к противнику рассматривались три варианта по убыванию меры агрессивности: подавление, доминирование или паритет, где в целевой функции используется информация об уровнях безопасности обеих систем. Допускалось динамическое переключение целей, если текущая стратегия не обеспечивала выживание своей системы. Алгоритмы управления в зависимости от постановки конкретной исследовательской задачи могут использоваться разные.

В предлагаемом докладе в развитие предшествующих исследований поставлена задача поиска и построения механизма, позволяющего воздействовать на стороны таким образом, чтобы объективно принуждать (может быть, угрозой ущерба) к отказу от агрессивных стратегий.

Динамическая модель с ресурсообменом

Очередной этап развития ДМП с включением в неё процесса ресурсообмена сторон является обобщением прежнего варианта модели, который становится частным случаем. Не изменяя прежнего состояния структурной модели и системы дифференциальных уравнений, добавляются новая связь и переменная ресурса $s^{(y)}$ и $s^{(z)}$.

Постановка задачи. Динамическая модель противоборства в условиях ресурсообмена представляется следующей структурной схемой (рис. 1).

На схеме показаны основные связи и элементы подсистем Y и Z, участвующие в ресурсообмене, а именно *ресурс* ($s^{(y)}$, $s^{(z)}$), *блок управления*, распределяющий ресурсы, *остальные элементы системы* (их полное описание дано в работе [10]), являющиеся фазовыми переменными модели: обеспечивающая деятельность y_1, z_1, система обеспечения безопасности (СОБ) y_2, z_2, целевая деятельность y_3, z_3, уровень угроз y_4, z_4, уровень безопасности y_5, z_5.

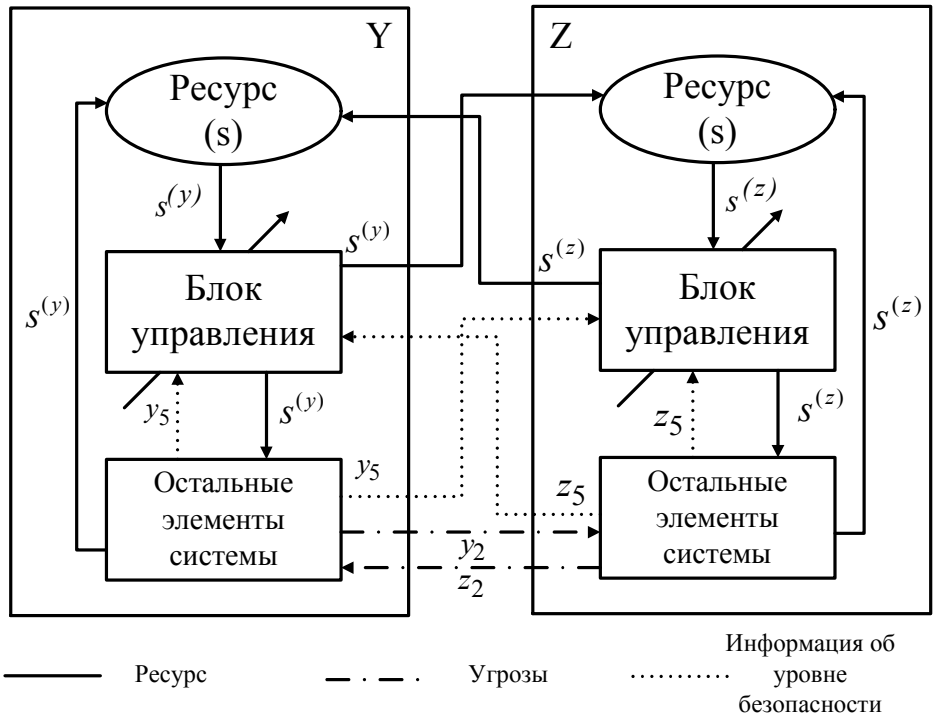


Рис. 1. Структурная схема динамической модели противоборства в условиях противоборства

Изменение состояния фазовых переменных каждой из подсистем описывается системой линейный дифференциальных уравнений M :

$$M = \begin{cases} Y\begin{cases} \dot{y}_i = f(t, \bar{y}, \dot{\bar{y}}, z_2, s^{(y)}) \\ s^{(y)} = f(t, s^{(y)}, y_3, s^{(z)}) \end{cases}, \ i = \overline{1..5}, \\ Z\begin{cases} \dot{z}_i = f(t, \bar{z}, \dot{\bar{z}}, y_2, s^{(z)}) \\ s^{(z)} = f(t, s^{(z)}, z_3, s^{(y)}) \end{cases} \end{cases} \quad (1)$$

где $\dot{y}$, $\dot{z}$ – производные от фазовых переменных.

В свою очередь уравнения ресурса принимают вид:

$$\begin{aligned} \dot{s}^{(y)} &= -k^{(y)} s^{(y)} + g^{(y)} y_3 + r_4^z s^{(z)} - r_4^y s^{(y)} \\ \dot{s}^{(z)} &= -k^{(z)} s^{(z)} + g^{(z)} z_3 - r_4^z s^{(z)} + r_4^y s^{(y)} \end{aligned},$$

где $g^{(y)}, g^{(z)}$ – коэффициент преобразования продуктов целевой деятельности; $k^{(y)}, k^{(z)}$ – коэффициенты, определяющие скорость расходования, r_4^z, r_4^y – доля, участвующая в обмене.

Распределением ресурса в подсистеме занимается блок управления, реализующий одну из стратегий подавление, доминирование или паритет, описывающиеся целевыми функциями:

$$\begin{aligned} F_{\text{пар}}(t) &= \left| y_5(t) - z_5(t) \right| < \delta, \\ F_{\text{под}}(t) &= y_5(t) - z_5(t) > \delta, \\ F_{\text{дом}}(t) &= z_5(t) < \delta, \end{aligned} \quad (2)$$

где $\delta \geq 0$ – параметр управления.

Характер взаимодействия подсистем в ходе противоборства. Проведенные на модели эксперименты, представленные в работе [8], показали, что при прочих равных условиях исход противоборства зависит от выбранной целевой функции блока управления. Паритет является миротворческой стратегией, где делается акцент на развитие целевой деятельности; доминирование и подавление более агрессивные стратегии, добивающиеся желаемого результата за счет увеличения влияния на противника.

На рисунках 2, 3 и 4 показаны графики изменения ресурса по ходу противоборства при ресурсообмене (линии черного цвета) и без него (линии серого цвета).

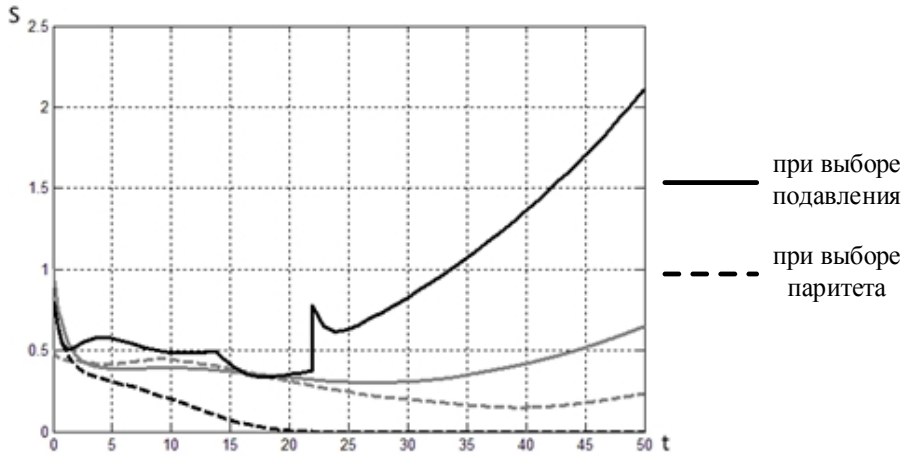


Рис. 2. Изменение ресурса подсистем с целевыми функциями подавления и паритета

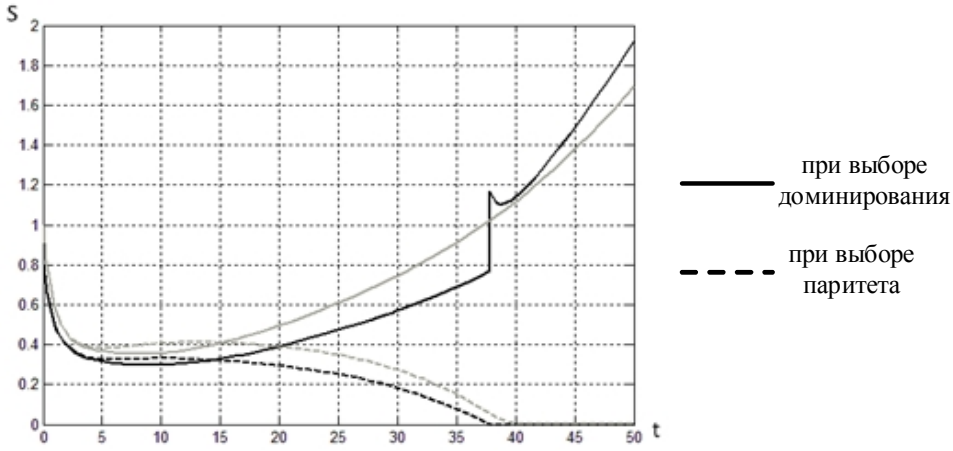


Рис. 3. Изменение ресурса подсистем с целевыми функциями доминирования и паритета

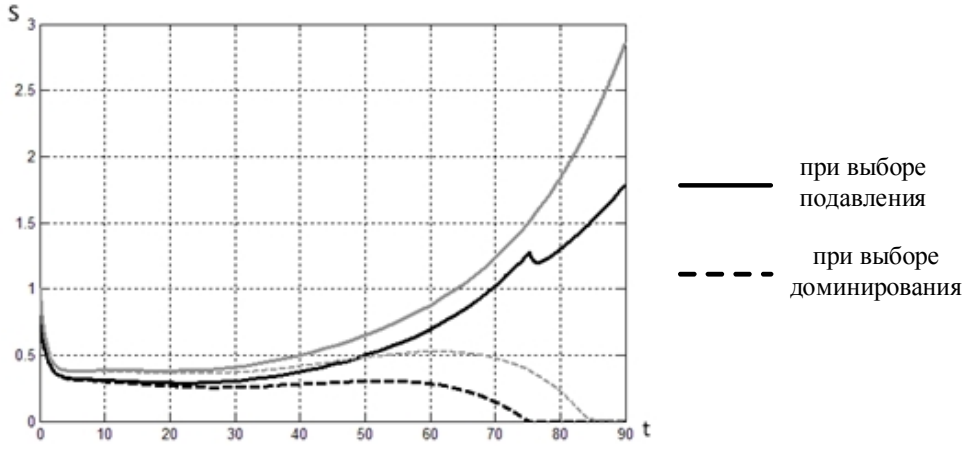


Рис. 4. Изменение ресурса подсистем с целевыми функциями подавления и доминирования

Стороны, выбравшие менее агрессивные стратегии, со временем теряют весь свой ресурс. Становится очевидным, что для выживания, участникам противоборства необходимо увеличивать своё влияние на противника. Для этого ресурсы внутри подсистемы перераспределяются от обеспечения целевой деятельности в сторону системы обеспечения безопасности. В резуль-

тате происходит наращивание атакующей и оборонительной способностей за счет снижения скорости развития своей основной деятельности.

В работе [8] дано более подробное представление хода противоборства, где исследовалось поведение сторон, выбравших агрессивные стратегии. На рисунке 5 представлены два варианта противоборства: «Паритет – Подавление» (сверху) и «Паритет – Доминирование» (снизу). У каждой из подсистем одинаковое начальное состояние (при $t = 0$), но разный ход развития. Серой линией изображен график стороны со стратегией паритет, черной – со стратегиями подавление и доминирование.

Подсистема со стратегией подавление (являющаяся самой агрессивной из стратегий) создает комфортное преимущество за счет развития своей системы обеспечения безопасности. Это происходит вплоть до уничтожения противника, после чего перераспределяет ресурс в сторону целевой деятельности. В случае выбора стратегии доминирование, преимущество достигается за счет недостаточной поддержки целевой деятельности, с развитием системы обеспечения безопасности.

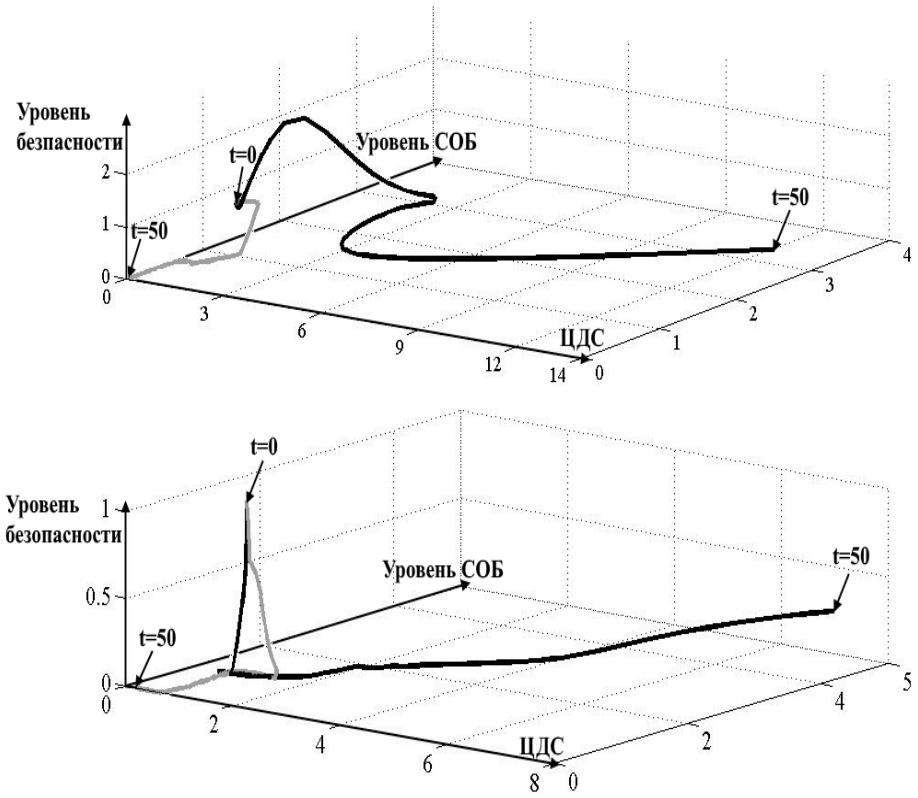


Рис. 5. Графики зависимости уровня безопасности, уровня системы обеспечения безопасности (СОБ) и целевой деятельности системы (ЦДС)

Получается, что агрессивность подсистемы характеризуется развитием системы обеспечения безопасности при деградации целевой деятельности.

Эксперименты на модели показали, что для сохранения приемлемого уровня безопасности для себя необходимо следовать стратегии не менее агрессивной, чем у противника. Такая тенденция вызовет рост взаимной агрессии систем, что может привести к деградации целевой деятельности, а, следовательно, истощению ресурса.

Модель гипервизора. Чтобы поддерживать уровень взаимной агрессии в допустимых пределах, предлагается ввести в модель самостоятельную алгоритмическую надстройку – «гипервизор», который будет получать информацию об участниках противоборства и генерировать не контролируемые ими управляющее воздействия, в целях понижения агрессии. Реализация подобного механизма пока не ясна, но стоит определить необходимые условия его существования.

Гипервизор должен оценивать ситуацию в разном масштабе времени. Для этого в его алгоритме присутствует переменная периодичности τ.

Сформируем критерии работы гипервизора (на примере подсистемы Y):

критерий контроля изменения системы обеспечения безопасности $\frac{\partial y_2}{\partial \tau} \le \theta_1$, где θ_1 – допустимое значение скорости развития СОБ;

критерий контроля изменения целевой деятельности $\frac{\partial y_3}{\partial \tau} \ge \theta_2$, где θ_2 – допустимое значение скорости развития целевой деятельности.

Уровень агрессии подсистемы будет считаться растущим, если нарушаются оба критерия вместе.

Для выполнения своих функций гипервизор должен удовлетворять следующим свойствам:

независимость – подсистемы Y и Z не могут влиять на гипервизор;

реактивность – управляющие воздействия подаются только в случае определения отклонения (т.е. когда замечается, что уровень агрессии одной из подсистем начинает расти);

оперативность - гипервизор должен выдавать своевременные управляющие воздействия подсистемам, пока ситуация не изменилась.

В свою очередь присутствие гипервизора, как активного наблюдателя сказывается и на поведении подсистем. Учитывая, что увеличение агрессии может повлечь за собой со стороны гипервизора стимулирующие меры разной направленности, подсистема может самостоятельно регулировать свою активность.

На рисунке 6 показана схема взаимодействия подсистем модели противоборства с гипервизором.

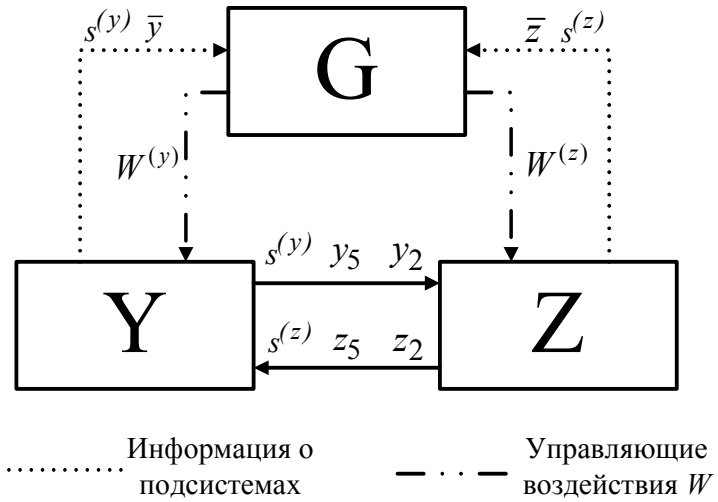


Рис. 6. Схема взаимодействия подсистем модели противоборства с гипервизором

Что касается управляющих воздействий, то они должны быть адаптивными, а это означает, что гипервизору необходимо оценивать изменение обстановки и менять со временем алгоритмы своей работы на более актуальные. Основные направления воздействия гипервизора можно разделить на две группы: алгоритм блока управления подсистемы и её элементы.

Влияя на блок управления подсистемы, можно побуждать к изменению параметра управления, из-за чего удовлетворение целевой функции будет достигаться при менее жестких требованиях; полностью изменять стратегию поведения, что уменьшит влияние на противника; устанавливать ограничения на распределение ресурсов, тем самым замедляя развитие атакующей способности.

Второй способ влияния – на элементы подсистемы, является более гибким и сложным, так как элементов в уравнении подсистемы достаточно много и все они взаимосвязаны. Среди них можно выделить наиболее значимые: система обеспечения безопасности y_2, z_2, уровень угроз

y_4, z_4, коэффициент влияния СОБ на уровень угроз противника p^z_{24}, p^y_{24}. Тогда для регулирования влияния подсистем друг на друга необходимо понижать y_2, z_2, p^z_{24}, p^y_{24} и повышать y_4, z_4 у более агрессивной стороны.

Третьим способом является комбинация первых двух, что только увеличивает гибкость в действиях гипервизора.

В общем виде, введение гипервизора в модель изменит систему дифференциальных уравнений (1) и целевые функции (2) следующим образом:

$$M = \begin{cases} Y\begin{cases} \dot{y}_i = f(t, \overline{y(W^{(y)})}, \bar{\dot{y}}, z_2, s^{(y)}) \\ s^{(y)} = f(t, s^{(y)}(W^{(y)}), y_3, s^{(z)}) \end{cases} \\ Z\begin{cases} \dot{z}_i = f(t, \overline{z(W^{(z)})}, \bar{\dot{z}}, y_2, s^{(z)}) \\ s^{(z)} = f(t, s^{(z)}(W^{(z)}), z_3, s^{(y)}) \end{cases} \end{cases}, \; i = \overline{1..5},$$

$$F_{\text{пар}}(t) = |y_5(t) - z_5(t)| < \delta(W^{(*)}),$$

$$F_{\text{под}}(t) = y_5(t) - z_5(t) > \delta(W^{(*)}),$$

$$F_{\text{дом}}(t) = z_5(t) < \delta(W^{(*)}),$$

где запись $y(W^{(y)})$, $s^{(y)}(W^{(y)}$, $z(W^{(z)})$, $s^{(z)}(W^{(z)})$ – означает зависимость фазовых переменных от влияния гипервизора; $\delta(W^{(*)})$ – зависимость параметра управления от управляющего воздействия гипервизора для стороны $(*) - Y$ или Z.

Предложенная реализация является не единственной и требует экспериментальной проверки на модели.

Заключение

В докладе рассмотрен вопрос управления конфронтационным взаимодействием с применением динамической модели противоборства. Предложенный подход рассматривает противоборство в общем случае и может применяться для систем разного рода и масштаба, от чисто технического уровня (противоборство информационной системы и злоумышленника через информационный канал) до более или менее глобальных разнородных по составу объектов.

Результаты предыдущих работ показали, что для выживания в условиях противоборства подсистемам необходимо переходить к более активным (агрессивным) стратегиям управления, за счет чего замедляется развитие целевой деятельности системы.

Для регулирования нарастающей агрессии предлагается включить в модель алгоритмический механизм – гипервизор, который бы приводил стороны, если не к гармонии, то хотя бы к долговременному, пусть даже вынужденному, паритету на приемлемом уровне противостояния.

Представлен первый вариант гипервизора, который дальше может дополниться, но сейчас важно было определить необходимые условия его существования, поэтому вопросы реализации остаются на перспективу.

Не менее интересно включить в систему третьего независимого участника взаимодействия, наличие которого, возможно, способно привести к желаемому результату, и механизм гипервизора станет имманентным самой системе.

Надо отметить, что напрашивающаяся аналогия наших выводов с реальными проблемами глобального взаимодействия в мире никоим образом априорно не закладывалась в модель, которая создавалась исключительно из общих абстрактных соображений.

По мере усложнения каждая версия модели открывала и продолжает открывать различные варианты направления исследований и экспериментов, но, наряду с поиском общих закономерностей поведения модели, их нужно актуализировать, вводя в конструктивное русло решения конкретных содержательно сформулированных задач.

ЛИТЕРАТУРА

1. Макаренко С.И. Информационное противоборство и радиоэлектронная борьба в сетецентрических войнах начала XXI века. Монография. — СПб.: Наукоемкие технологии, 2017. — 546 с.
2. Макаренко С.И., Михайлов Р.Л. Информационные конфликты – анализ работ и методологии исследования // Системы управления, связи и безопасности. 2016. № 3. С. 95-178.
3. Губанов Д.А., Новиков Д.А., Чхартишвили А.Г. Социальные сети: модели информационного влияния, управления и противоборства. М.: Физматлит, 2010. 244 с.
4. Расторгуев С.П. Информационная война. Проблемы и модели. Экзистенциальная математика: учебное пособие для студентов, обучающихся по специальности в области информационной безопасности / С.П. Расторгуев — М.: Гелиос АРВ, 2006. — 240 с.
5. Расторгуев С.П., Токарев Р.С. О перспективах глобальной технической среды информационного противоборства // Информационные войны. 2009. 2(10). С.9-15.
6. Юсупов Р.М., Шишкин В.М. Информационно-коммуникационные технологи и национальная безопасность - противоречивая реальность // Информатизация и связь, № 1, 2010. С. 27-35.
7. Колесников К.Е., Шишкин В.М. Исследование динамики симметричного противоборства на дифференциальной модели // Информационная безопасность социально-экономических систем: монография / Апатова Н.В. и др. / под ред. проф. О.В. Бойченко. Симферополь: ИП Зуева Т.В., 2017. С. 202-218.
8. Шишкин В.М., Колесников К.Е. Динамическая модель противоборства – интерпретации и эксперименты // Анализ, моделирование, управление, развитие социально-экономических систем: сборник научных трудов XI Международной школы-симпозиума АМУР-2017, Симферополь-Судак, 14-27 сентября 2017 / под общ. ред. А.В. Сигала. Симферополь: ИП Корниенко А.А., 2017. С. 438-444.
9. Шишкин В.М. Социально-экономическое развитие в контексте национальной безопасности и научно-технического развития. Динамическая модель // Анализ, моделирование, управление, развитие экономических систем: сборник научных трудов VIII Международной школы-симпозиума АМУР-2014, Севастополь, 12-21 сентября 2014 / Под ред. доцента А.В. Сигала. Симферополь: ТНУ им. В.И. Вернадского, 2014. С. 359-364.
10. Шишкин В.М., Колесников К.Е. Исследование процессов противоборства средствами имитационного моделирования // Труды Восьмой всероссийской научно-практической конференции «Имитационное моделирование. Теория и практика» (ИММОД-2017) (г. Санкт-Петербург, 18-20 октября 2017 г.). СПб.: Изд-во ВВМ, 2017. С. 213-218.

V.M. Shishkin (SPIIRAS, Saint-Petersburg), **K.E. Kolesnikov** (Main department of Ministry of justice in Russia in St. Petersburg, Saint-Petersburg)

Control confrontation interaction in the conditions of resource exchange - a dynamic model

Interaction in the modern world has considerably gained confrontational character irrespective of the scale of the considered objects, at the same time the complexity of systems, heterogeneity, but, at the same time, integration of their elements have increased that leads to emergence of a set of the various communications (sociotechnical, cyberphysical, etc.) complicating understanding and the description of the processes proceeding in them, especially, mathematically correct, but rather compact.

The purpose of the represented work is the research of a possibility of application of dynamic modeling, the device of the differential equations, for the description of processes in badly formalizable contradictory systems and forecasting of their behavior for various scenarios, criteria and control algorithms.

The model is based on earlier researches of interaction of development of information and communication technologies and the ensuring national security, which has shown plausible and useful results, and her further development already as confrontation models at preservation of the basic functional elements.

The model is scaled, and without change of structure if to enclose a bit different substantial sense in some phase variables, it can be interpreted in the wide range from interstate cooperation to, for example, corporate or information technological level. For example the possibility of her integration with the existing control systems of the events of network safety functioning in the monitoring mode has been shown.

Each of the parties of confrontation has the control unit which main objective is distribution of a resource in system. He realizes three essentially different strategy of management called "parity", "suppression" and "domination" which combination formally creates nine options of interaction of the parties.

Originally the parties were considered as autonomous systems which interaction was limited only to management of threats to security concerning the opponent. In the considered model one more channel of interaction through resource exchange, doing the parties not only confronting, but also interdependent is added that changes dynamics of variables and a possibility of management of process, allowing to realize, in particular, a superstructure in the form of the forcing regulator hypervisor.

The new channel creates two options of interaction: with one-sided and with two-sided exchange of resources. Considering earlier defined strategy of management, eighteen new options of development of the relations between systems turn out. Decrease in level of aggression between the dependent parties was additional effect of introduction of a hypervisor that in turn favorably affects the level of safety and development of systems separately and accelerates their development. Nevertheless, the algorithm of work of a hypervisor plays a key role therefore his improvement represents an independent task.

The made computing experiments and the received new results, as well as in the previous modifications of model, have shown her plausible behavior that gives the grounds and opportunities for use of this model at the solution of the applied tasks connected with complex confrontation including, in the information sphere at the different levels of community.

МЕТОДИКА ОБОСНОВАНИЯ ПАРАМЕТРОВ ДИСЦИПЛИНЫ ОБСЛУЖИВАНИЯ РАЗНОРОДНЫХ АБОНЕНТОВ МУЛЬТИСЕРВИСНОЙ СЕТИ ДОСТУПА ПРИ ЗАДАННЫХ ТРЕБОВАНИЯХ К ХАРАКТЕРИСТИКАМ ИНФОРМАЦИОННОГО ОБМЕНА

С. Е. Потапов
(*Филиал Военной академии РВСН имени Петра Великого в г. Серпухове, Московская обл. г. Серпухов, ул. Бригадная д. 17, 8(909)164-71-60*)

Аннотация
В данной работе рассматривается подход к оптимизации настройки параметров обслуживания мультисервисных абонентов в сетях доступа для повышения эффективности использования коммуникационных ресурсов при обеспечении заданных требований по качеству информационного обмена между абонентами с различным типом трафика

Ключевые слова: Сети доступа, система массового обслуживания, дисциплина обслуживания.

Введение

В настоящее время в различных сферах человеческой деятельности нашли широкое применение информационно-телекоммуникационные сети гибридной конфигурации. Распределённые информационные сети обычно строятся по принципу организации очаговых сетей доступа, имеющих выход в высокоскоростные магистральные каналы связи. При этом, в условиях ограниченности коммуникационного ресурса каналов доступа к магистральным сетям, зачастую, нерационально организованы дисциплины обслуживания абонентских информационных потоков, так как не учитывают задаваемые административные требования к тем или иным видам предоставляемых телекоммуникационных сервисов связи. Поэтому актуальной является задача разработки научно-обоснованного методического аппарата обоснования параметров обслуживания абонентов различного типа в сетях доступа с заданными требованиями по качеству информационного обмена.

С появлением мультисервисных абонентов связи [1], требующих для качественного информационного обмена обеспечения достаточно жёстких и порой противоречивых требований к характеристикам функционирования транспортных сетей, задача эффективного распределения коммуникационного ресурса сетей доступа приобрела новые очертания. Жесткое закрепление необходимого количества пропускной способности каналов доступа за разными типами абонентов позволяет обеспечить в любой момент времени требуемые характеристики информационного обмена, однако с экономической точки зрения крайне невыгодно. Случайный доступ к общей пропускной способности сети доступа позволяет существенно сэкономить коммуникационный ресурс оператора связи, но создаёт риски отказа абонентам в предоставлении каналов связи по причине их занятости. Учитывая различные требования сервисов к задержке пакетов и вероятности их потери, а также наличие в зоне обслуживания сети доступа приоритетных абонентов задача эффективного управления коммуникационным ресурсом сводится к определению оптимального размера емкости общего канала доступа и распределение приоритетов между абонентами.

Под критерием эффективности функционирования сети доступа будем понимать комплексный показатель Y, включающий степень выполнения требований по вероятности отказа и времени предоставления доступа, а также величину затрачиваемого общего коммуникационного ресурса:

$$Y =< E, T, W > \quad (1)$$

При этом, для различных m типов абонентов

$$E = \sum_{j=1}^{m} I_j \cdot k_j , \quad (2)$$

где $j = [1,m]$ – частные вероятности отказа абонентского трафика, k_j – их степени важности ($\Sigma k_j = 1$), а I_j – переменная (индикатор), принимающая значение «1» – если $\text{Ротк}_j \le \text{Ротк(доп)}_j$ и «0» – в обратном случае.

Если существуют критически важные сервисы абонентов, то выражение (2) примет вид:

$$E = I_{кр}\left(k_{кр} + \sum_{j=1}^{m} I_j \cdot k_j\right), \tag{3}$$

где $I_{кр} = \prod_i I_{кр}^i$ - произведение индикаторов критических сервисов $k_{кр}$ – коэффициент их важности.

При рассмотрении сетей доступа без ожидания соединения предполагается, что в случае наличия свободного коммуникационного ресурса доступ к нему предоставляется мгновенно, в обратном случае – абоненты получают отказ в обслуживании. Таким образом, эффективность функционирования таких сетей определяется степенью выполнения требований по вероятности отказа и объемом затрачиваемых на это ресурсов.

Для оценивания эффективности предполагаемой конфигурации сети доступа необходимо рассматривать стохастический показатель качества её функционирования, который выражается вероятностью выполнения требований к информационному обмену и сохранением заданного уровня ресурсоёмкости:

$$P(\tilde{Y} \in Y^{доп}) = P\left((\tilde{E} \ge E^{доп}) \cap (\tilde{W} \le W_{доп}\right) \tag{4}$$

$$P(\tilde{E} \ge E^{доп}) = P\left(I_{кр}\left(k_{кр} + \sum_{j=1}^{m} I_j \cdot k_j\right) \ge E^{доп}\right) = F_E(E = E^{доп}), \tag{5}$$

где F_E – функция распределения показателя E, имеющая следующий вид:

$$F(E) = \begin{cases} 0: & E < 0; \\ 1 - P(I_{кр}): & E \in [0; k_{кр}); \\ \gamma(\bar{k}; \bar{I}; m): & E \in [k_{кр}; 1]; \\ 1: & E > 1. \end{cases}, \tag{6}$$

где, в свою очередь $\gamma(\bar{k}; \bar{I}; m) = rang(\chi(\phi)); \quad \chi_i(\phi_i) = \prod_{j=0}^{n1} \bar{P}(I_j) \cdot \prod_{j=n1+1}^{n2} \bar{P}(I_j); \quad \phi_i = \sum_{j=n1+1}^{n2} k_j$;

$n_1, n_2 = [1,m], \; n + n_2 = m$

На рисунке 1 представлена обобщённая схема функционирования сети доступа с мультисервисными абонентами.

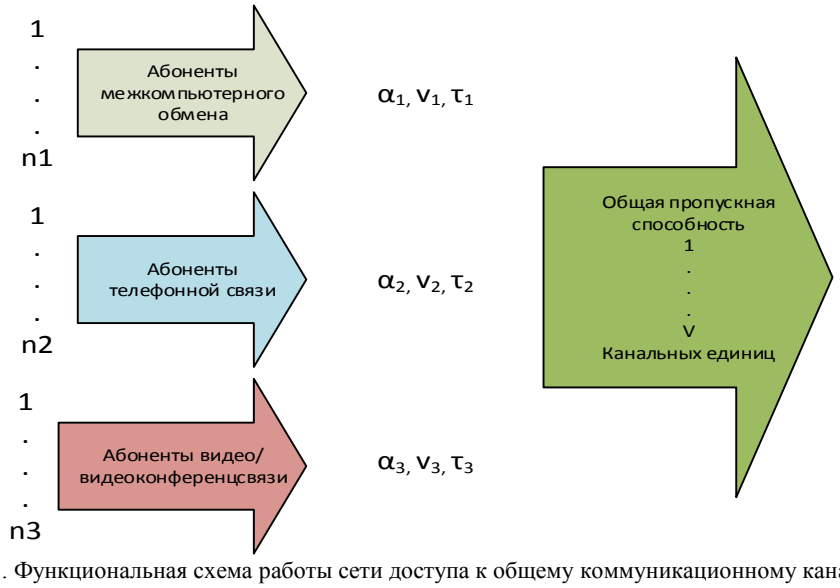


Рис. 1. Функциональная схема работы сети доступа к общему коммуникационному каналу

Здесь представлены три группы по ni абонентов использующих для работы различных сервисов связи общий канал доступа. При этом предполагается, что моменты поступления заявок на предоставление необходимого количества vi пропускной способности общего канала доступа распределены по пуассоновскому закону с параметром αi, а также известно среднее время занятия этой пропускной способности τi. Такой процесс функционирования сети доступа хорошо описывается математическим аппаратом конечных марковских цепей как мультисервисная модель замкнутой системы массового обслуживания (СМО) с потерями [2]. Граф переходов состояний такой цепи для двух типов абонентов представлен на рисунке 2.

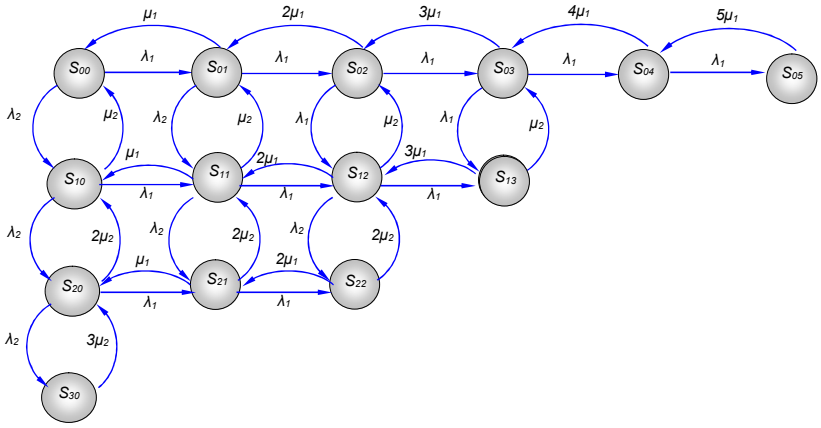


Рис. 2. Пример КМЦ для двух типов неприоритетных абонентов

Математическое описание такой модели достаточно полно приведено в [1]. В частности, представлены рекуррентные соотношения, позволяющие быстро получить требуемое количество канальных единиц, обеспечивающие заданные требования по вероятности отказа для всех типов абонентов (одна канальная единица есть наибольший общий делитель требуемой пропускной способности всех типов абонентов сети доступа). Очевидно, что определяющим фактором здесь является обеспечение заданного уровня вероятности отказа для самых требовательных в этом плане абонентов. При этом для остальных абонентов данная характеристика будет завышена, что влечёт за собой нерациональное использование коммуникационного ресурса.

На рисунке 3 представлены результаты исследования зависимости вероятности отказа в предоставлении требуемого количества канальных ёмкостей для совместной работы двух типов абонентов при следующих исходных данных: количество абонентов обоих типов n1 = n2 = 20; интенсивность запроса на соединение λ1 = 0,2 λ2 = 0,3 с-1; интенсивность освобождения канала μ1 = μ1 = 1/τ = 2 с-1; требуемая канальная ёмкость v_1 = 2 v_2 = 3 канальных единицы; общая пропускная способность канала доступа изменялась от 10 до 26 канальных единиц.

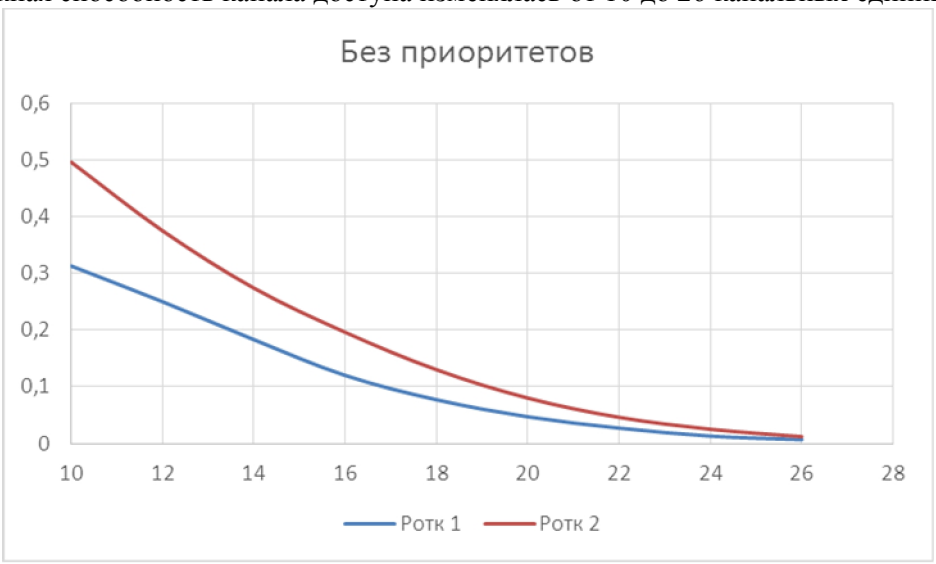


Рис. 3. Графики зависимости вероятности отказа от общей пропускной способности канала доступа

Из анализа графиков следует, что для обеспечения заданного уровня вероятности отказа необходимо ориентироваться на поведение второго типа абонентов, вероятность отказа у которых выше. Например, для обеспечения вероятности отказа Ротк = 0,1 необходимо предоставить общий канал доступа 19 канальных единиц.

Повысить эффективность функционирования такой системы возможно, если предоставить самым требовательным к вероятности отказа абонентам приоритет доступа к общему каналу. Тогда модель функционирования сети доступа будет описываться следующим графом переходов состояний процесса (рисунок 4).

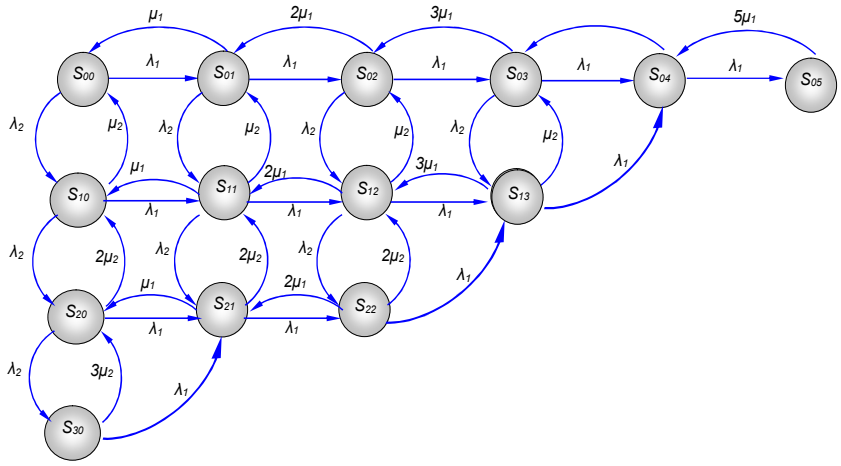


Рис. 4. Пример КМЦ для двух типов абонентов с приоритетами

Наличие дополнительных переходов, связанных с приоритетным обслуживанием наиболее требовательных абонентов, затрудняет получение рекуррентных соотношений для вероятностей отказа каждой группы мультисервисных абонентов. Поэтому исследование таких моделей необходимо производить по общему алгоритму с построением системы уравнений информационной баланса нагрузки для стационарного режима [3].

Анализ закономерностей построения графов переходов состояний в таких моделях позволил получить правила автоматизированного синтеза основной матрицы системы уравнений баланса для общего случая:

- для каждого типа абонентов переходы в сторону возрастания нумерации состояний определяются выражением (4), а переходы в сторону убывания нумерации состояний – выражением (5);

$$a_{(i,j)(i+1,j)} = (n_i - i)\lambda_i \tag{4}$$

$$a_{(i,j)(i-1,j)} = i\mu_i \tag{5}$$

- элементы главной диагонали матрицы А определяются по выражению (6);

$$a_{(j,j)} = -\sum_i a_{ij}, \quad (i \neq j) \tag{6}$$

- определяются состояния отказа приоритетных абонентов ($S_{ij} \in U$) по следующему правилу:

$$3i + 2(j+1) > 10 \tag{7}$$

- из каждого состояния отказа, за исключением крайнего, осуществляется переход в состояние $S_{i,(j+1)}$ с интенсивностью $(n_j\text{-}j)\lambda_2$;

Решается система линейных алгебраических уравнений относительно вероятностей состояний системы, например, матричным методом.

На рисунке 5 представлены результаты исследования зависимости вероятности отказа в предоставлении требуемого количества канальных ёмкостей в условиях предыдущего примера для случая, когда вторая группа абонентов имеет абсолютный приоритет доступа к общему каналу перед абонентами первой группы.

Из анализа графиков следует, что за счет введения приоритетов в обслуживании график зависимости вероятности отказа второй группы абонентов опустился ниже графика первой груп-

пы, однако в данном случае для обеспечения требуемой вероятности отказа Ротк = 0,1 необходимо предоставить общий канал доступа уже в 18 канальных единиц.

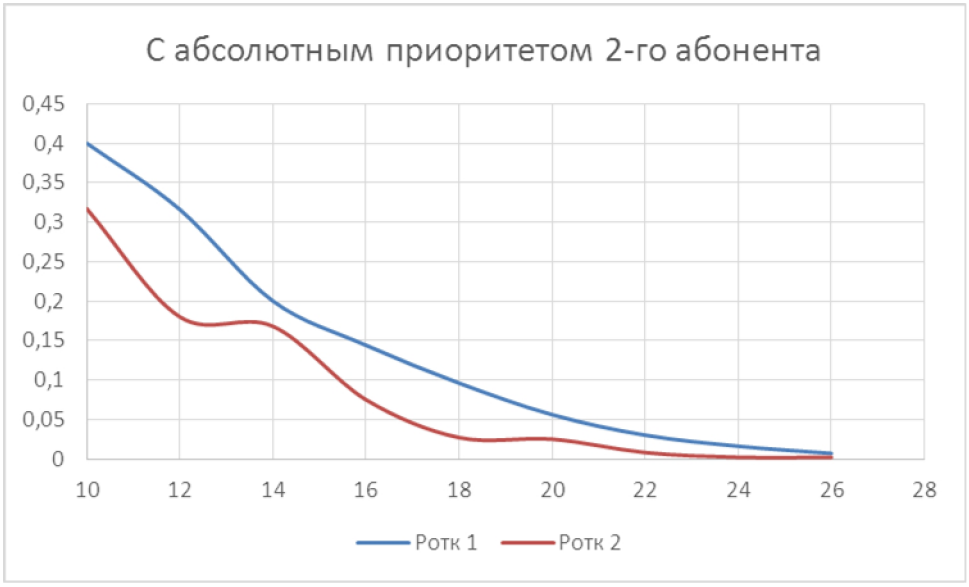


Рис. 5. Графики зависимости вероятности отказа от общей пропускной способности канала доступа с абсолютными приоритетами

Дальнейшее увеличение эффективности использования общей пропускной способности канала доступа можно добиться за счёт сближения графиков зависимости вероятности отказа. Для этого необходимо ввести относительные приоритеты доступа к каналу наиболее требовательным абонентам. При этом коэффициент относительности К определяет долю получающих приоритет заявок на соединение в общем потоке заявок приоритетных абонентов. Тогда в модели СМО интенсивности переходов из предельных состояний, соответствующие поступлению приоритетных заявок необходимо умножить на коэффициент К, а вероятности отказа приоритетных абонентов дополнить вероятностями пребывания процесса в этих состояниях с коэффициентом (1-К). Графики зависимости вероятностей отказа от ёмкости канала доступа в этом случае при К = 0,8 представлены на рисунке 6.

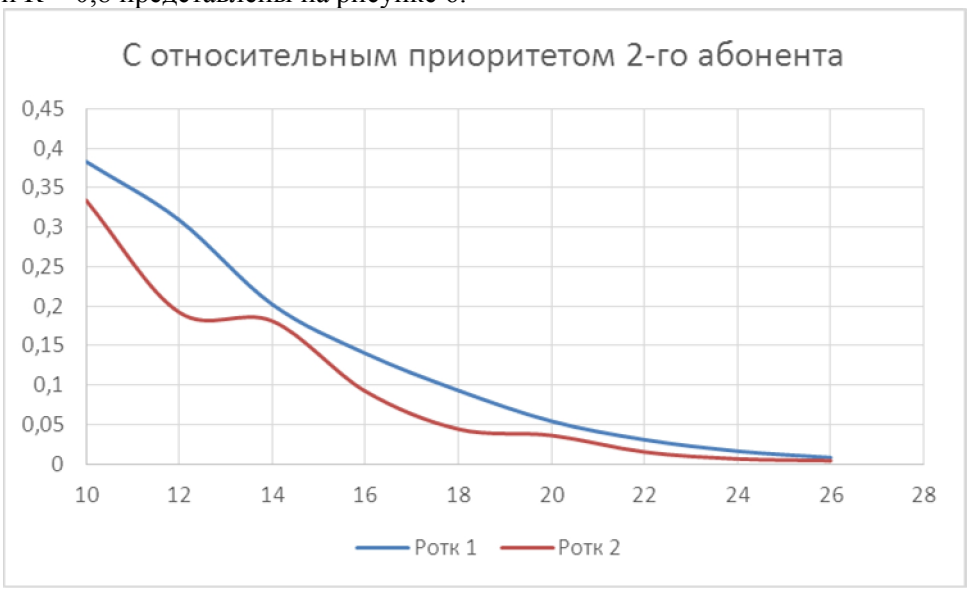


Рис. 6. Графики зависимости вероятности отказа от общей пропускной способности канала доступа с относительными приоритетами

Сближение графиков зависимостей вероятности отказа различных типов абонентов обуславливает скорейшее выполнение требований при меньшем объеме пропускной способности общего канала доступа.

Таким образом, методика обоснования параметров дисциплины обслуживания разнородных абонентов мультисервисной сети доступа при заданных требованиях к характеристикам информационного обмена сводится к выполнению следующих операций:

1. Определение допустимого значения вероятностей отказа для различных типов абонентов и допустимой ёмкости общего канала доступа (в канальных единицах).
2. Формирование критерия показателя эффективности в виде (4) – (6).
3. Построение математической модели функционирования сети доступа в виде замкнутой СМО с относительными приоритетами.
4. Последовательное приближение функциональных зависимостей вероятности отказа от ёмкости канала доступа к их допустимым значениям с помощью подбора коэффициентов относительности приоритетов.
5. Исследование обобщённого показателя эффективности функционирования сети доступа на экстремум.
6. Вывод результатов исследования – получения оптимального значения ёмкости общего канала доступа.

Данный научно-методический аппарат позволит адаптировать дисциплину обслуживания мультисервисных абонентов под изменяющиеся условия информационной нагрузки и качество каналов связи. Информацию для прогнозирования данных параметров можно получить, например, как описано в [4].

ЛИТЕРАТУРА

1. Степанов С.Н. Теория телетрафика: концепции, модели, приложения. – М.: Горячая линия – Телеком, 2015. 868 с: ил. – (Серия «Теория и практика инфокоммуникаций»).
2. Кемени Джон Дж., Снелл Дж. Ларк. Конечные цепи Маркова./Пер. с англ. – М.: Наука, 1970.
3. Потапов С.Е., Тоискин В.Е., Якимова И.А., Рублёв В.Д., Солохин А.А. О функционировании коммутационного оборудования системы связи с незакреплёнными каналами / REDS: Телекоммуникационные устройства и системы. 2015. Т. 5. № 4. С. 360-363.
4. Потапов С.Е. Математическая модель статистической обработки информации о состоянии сети / В сборнике: РАДИОЛОКАЦИЯ, НАВИГАЦИЯ, СВЯЗЬ XXII международная научно-техническая конференция. 2016. С. 500-507.

Секция 3
МУЛЬТИАГЕНТНЫЕ СИСТЕМЫ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И НЕЙРОТЕХНОЛОГИИ В ЗАДАЧАХ УПРАВЛЕНИЯ

ГРУППОВОЕ УПРАВЛЕНИЕ МУЛЬТИАГЕНТНОЙ СИСТЕМОЙ АВТОНОМНЫХ НЕОБИТАЕМЫХ ПОДВОДНЫХ АППАРАТОВ ПРИ ВЕДЕНИИ ВСЕПОГОДНОЙ СЕЙСМОРАЗВЕДКИ

О. В. Карсаев[1]
(*СПИИРАН, С.-Петербург, 14 линия 39, karsaev@ips-logistic.com*)

Л. А. Мартынова[2]
(*АО «Концерн «ЦНИИ «Электроприбор», С.-Петербург, ул. Малая Посадская 30, martynowa999@bk.ru*)

Аннотация

В статье описывается модель автономного командного поведения группы АНПА, выполняющей миссию сейсморазведки заданной территории морского шельфа. Модель поведения разрабатывается на основе мультиагентного подхода. Основным объектом исследований в статье является схема информационного взаимодействия. Целью информационного взаимодействия является обеспечение всех АНПА оперативной информацией о текущем состоянии группы. При этом полагается, что данная информация может использоваться агентами АНПА для автономного планирования последующих действий АНПА. Автономное планирование действий выполняется в соответствии с априори определяемыми схемами построения, передвижения и перестроения группы АНПА. Схемы перестроения рассматриваются в случаях выхода из строя каких-то АНПА.

Ключевые слова: сейсморазведка, группа АНПА, схема информационного взаимодействия.

Введение

В настоящее время существует широкий круг задач, связанных с необходимостью использования группы подводных аппаратов. Одной из таких задач является ведение сейсморазведки морского дна. Наличие группы аппаратов необходимо для того, чтобы обеспечить одновременный прием отраженных от морского дна сигналов с различных направлений.

Исследованию и разработке методов группового управления робототехническими комплексами в настоящее время посвящено значительное количество работ. Теоретические основы группового управления подводных аппаратов, в том числе при наличии препятствий, приведены в [1-4]. Групповое управление автономных аппаратов с использованием наиболее адекватной мультигентной технологии, в соответствии с которой группа аппаратов моделируется в виде мультиагентной системы, а каждый автономный аппарат соответствующим агентом, описано в [5]. На основе использования такого подхода в работе [6] рассматривается функционирование группы беспилотных летательных аппаратов, в работе [7] - организация взаимодействия космических аппаратов.

Использование описанных подходов в указанных работах может рассматриваться в качестве теоретического базиса для решения исследуемой в данной работе задачи. Но при этом возникает необходимость в изучении достаточно широкого круга вопросов, связанных с учетом существующих ограничений и специфики функционирования автономных необитаемых подводных аппаратов в морской среде, в частности:

- по точности автономной подводной навигации,
- по дальности и скорости звукоподводной связи,
- по энергопотреблению.

Частично ответы на указанные вопросы получены в рамках выполнения проекта WiMUST (Widely Scalable Mobile Underwater Sonar Technology) [8]. Этот проект, финансируемый Евросоюзом, направлен на разработку интеллектуального, управляемого, распределенного и рекон-

[1] Кандидат технических наук, старший научный сотрудник.
[2] Доктор технических наук, ведущий научный сотрудник.

фигурируемого подводного акустического массива, который мог бы значительно повысить эффективность методологий, используемых для выполнения геофизических и геотехнических акустических съемок в море. По замыслу проекта подводные аппараты тянут стримеры с несколькими датчиками на каждом вблизи поверхности дна. Положение датчиков образует антенную решетку, которая в процессе выполнения миссии в зависимости от получаемых данных и внешней обстановки может оперативно менять свою топологию, и тем самым обеспечивать наиболее эффективное выполнение сейсморазведки.

Одной из главных задач, решаемых в рамках проекта, является реконфигурация топологии положения подводных аппаратов в группе при выполнении миссии. В связи с этим в проекте проработан вопрос использования гибких алгоритмов планирования миссии группы роботов. Так, инновационным является подход к планированию миссии, заключающийся в расширении классических алгоритмов планирования миссии путем использования новых интеллектуальных методологий, основанных на информации и количественном расширении полномочий АНПА, зависящем от их текущего состояния [9,10]. Такой подход может быть использован для генерации предпочтительного поведения АНПА без традиционного применения функций затрат, которые обычно меняются от задачи к задаче и назначаются заранее оператором. С технической точки зрения расширение прав и возможностей АНПА при выборе маршрута движения определяется как пропускная способность канала Шеннона между элементами управления АНПА и его состоянием в последующие моменты времени. Расчет пропускной способности основан на адаптивной оценке взаимодействия АНПА группы с окружающей средой. Используются характеристики АНПА, встроенные для оценки желательных траекторий АНПА с точки зрения его мобильности и живучести. В результате при выполнении команды управления общий планировщик миссии может оценить полезность выбранного маршрута путем отслеживания текущего состояния каждого АНПА. Кроме того, осуществляется выявление АНПА, находящихся или приближающихся к нежелательному состоянию, характеризуемому отсутствием мобильности (застревание, поломка), поскольку в этом случае возможность выбора поведения отсутствует.

Еще одной задачей, решаемой в рамках проекта, является навигация АНПА. В настоящее время разработан подход к расчету кооперативного маршрута, построенного на решениях [11] и характеризуемого явным учетом потерь связи и временных задержек. Так, алгоритм кооперативного управления предполагает, что каждый АНПА имеет полное представление о маршруте, который должен быть преодолен в ближайшее время с привязкой ко времени, и что маршруты всех АНПА параметризуются с помощью координат вдоль пути.

На начальном этапе миссии, когда АНПА переходит к заданной путевой точке, чтобы достичь заданного положения используется одномаяковая навигационная система для улучшения эффективности локализации положения АНПА. Для этого используется новая стратегия, предложенная в [12], в которой рассматривается проблема нахождения АНПА в заданном положении при оптимизации производительности одномаяковой навигации. Подходы, предложенные в [13], также могут использоваться для решения задачи относительной навигации нескольких АНПА.

Навигация АНПА неотрывно связана с коммуникацией аппаратов между собой. Для этого в рамках проекта WiMUST разработаны алгоритмы и процедуры точного измерения расстояния (до сантиметровой точности) между АНПА, движущимися в группе, восстановление их относительных положений в трехмерном пространстве и точная синхронизация их локальных часов. Исследования направлены на развитие алгоритмов коммуникации. Использование широкополосной коммуникационной технологии EvoLogics [14] открывает путь к практически достижимым скоростям от нескольких бит до десятков килобит в секунду. Хотя диапазон передачи данных и мощности подводного акустического канала может сильно изменяться в результате изменений условий окружающей среды. Эффективная реализация сети акустической связи существенно зависит от возможности модема автоматически отрегулировать свою битовую скорость до фактических условий канала.

Все описанные результаты исследований группового применения АНПА могут быть использованы в качестве основы для дальнейших исследований по взаимодействию АНПА в условиях всепогодности и наличия ледового покрытия, связанных с отсутствием судна сопровождения, которое используется в проекте WiMUST.

В рассматриваемой задаче полагается, что к каждому АНПА прикреплен геофон, а движение группы АНПА синхронизировано с движением излучателя и может осуществляться по схеме, описанной в [15-20]. Связь между АНПА группы необходима для обеспечения:

- навигации;
- координации;
- перегруппировки в случае выхода одного или нескольких АНПА из строя.

Задачами, необходимыми для решения, в этом случае являются

- навигация каждого АНПА без судна сопровождения и других элементов, кроме АНПА и излучателя;
- координация действий в связи с выполнением АНПА разнородных функций;
- перегруппировка в случае выхода АНПА из строя с учетом разнородности выполняемых функций и позиционирования АНПА без участия судна сопровождения.

Эти задачи должны быть решены с использованием информационного взаимодействия, которое имеет особенности в морской среде, о которых говорилось выше. Аппараты могут общаться с использованием различных технических средств и различных схем.

Основной целью настоящей работы является разработка схемы информационного взаимодействия подводных аппаратов при их групповом использовании.

Для достижения данной цели в работе были решены следующие задачи:

- рассмотрены особенности группового функционирования АНПА на примере ведения сейсморазведки;
- рассмотрены основные способы взаимодействия аппаратов в группе для обеспечения эффективного функционирования группы при выполнении поставленной задачи;
- разработана схема информационного взаимодействия при решении задач позиционирования, координации выполняемых действий, перегруппировки в случае замены вышедших из строя аппаратов.

Схема информационного взаимодействия

1. Модель процесса сейсморазведки и модель поведения АНПА

Процесс сейсморазведки, выполняемый группой АНПА, состоит из трех этапов: начальный этап - выгрузка всех АНПА из кросс-дока и их переход в исходные точки плана, основной этап - проведение сейсморазведки определенной территории шельфа, и заключительный этап – возвращение и загрузка АНПА в кросс-док. Объектом исследований данной работы является только основной этап.

Основной этап может выполняться в соответствии с предопределенным планом сейсморазведки, основой которого является представление обследуемой территории шельфа в виде множества маршрутных точек. Целью сейсморазведки является сбор данных в этих точках: прием отражаемых морским шельфом сигналов, посылаемых излучателем. Модель выполнения данного плана, обоснованная в предшествующих исследованиях [15-16], кратко может быть представлена следующим образом. Группа АНПА образует строй прямоугольной формы с предопределенным расстоянием между соседними рядами и шеренгами. Сценарий обхода всех маршрутных точек представляется в виде множества последовательных итераций. В рамках каждой итерации n-1 рядов строя АНПА находятся в маршрутных точках плана, излучатель передвигается над этими АНПА по заданной траектории с определенной периодичностью посылает сигналы, а один ряд АНПА выполняет переход в последующие маршрутные точки плана сейсморазведки.

На основе описанной модели выполнения плана сейсморазведки в целом определяется модель автономного поведения каждого АНПА, в соответствии с которой каждый АНПА последовательно выполняет типовой сценарий поведения (рис. 1). Сценарий поведения предопределяет последовательное нахождение АНПА в одном из шести состояний

a) переход в следующую маршрутную точку плана,
b) заглубление и прикрепление ко дну,
c) ожидание начала сеанса сейсморазведки,
d) прием отраженных сигналов (сеанс сейсморазведки),

e) планирование следующей итерации,
f) всплытие на заданную глубину для дальнейшего передвижения,
и также предопределяет схему информационного взаимодействия АНПА с остальными аппаратами группы.

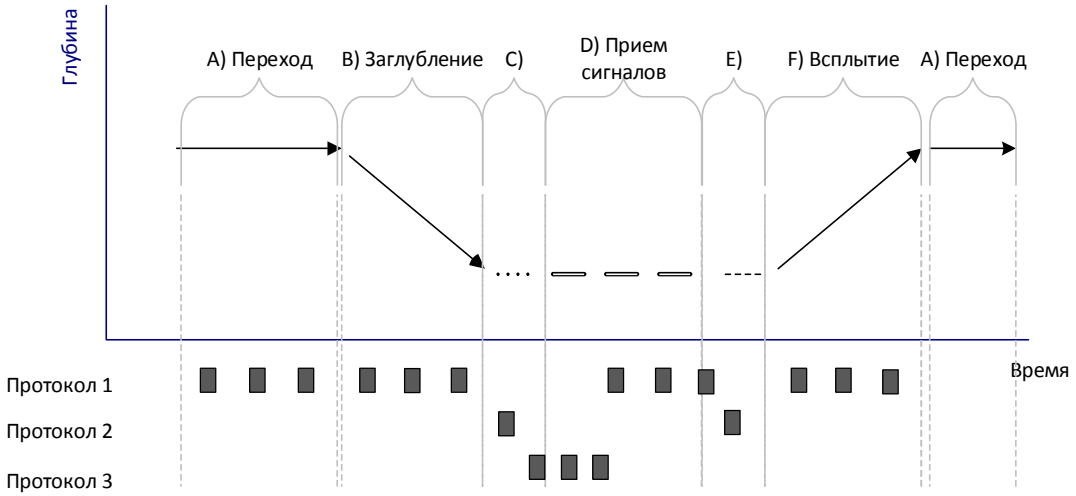


Рис. 1. Типовой сценарий поведения АНПА

Схема информационного взаимодействия в соответствии с рисунком 1 определяется тремя протоколами, и предназначается для решения следующих основополагающих задач автономного управления группировкой:

- определение текущих координат АНПА, находящихся в состоянии *a, b* или *f,* т.е. в состоянии движения (протокол 1),
- уточнение текущих координат АНПА, находящихся в состоянии *d,* т.е. прикрепленных ко дну (протокол 3),
- выявление ситуаций недопустимого сближения АНПА для избегания их столкновения (протокол 1)
- поддержание в оперативной памяти всех АНПА текущей осведомленности о состоянии и движении группировки в целом (протоколы 1 и 2).

Детальное описание протоколов взаимодействия и пояснения к решению соответствующих задач приводятся далее в пункте 3 после описания физических и технологических возможностей информационного взаимодействия в морской среде в пункте 2.

2. Физические и технические возможности информационного взаимодействия

В данном случае физические возможности и ограничения, главным образом, определяются тремя основными факторами.

Первый фактор. Информационное взаимодействие внутри группировки выполняется в рамках динамической телекоммуникационной сети, которая формируется всеми АНПА группировки. При этом технологические основы передачи сообщений в рамках такой сети, главным образом, определяются сопоставлением двух параметров: d^1 – максимально возможная удаленность между произвольной парой АНПА группировки в процессе выполнения миссии, и d^2 - дальность распространения гидроакустических сигналов, которая определяется устанавливаемыми на АНПА гидроакустическими модемами. Максимально возможная дальность действия гидроакустического модема в настоящее время ограничивается значением в 2000 метров. Описание данного модема можно найти в работе [21].

Если $d^1<d^2$, то в любой момент времени между любой парой АНПА в группировке возможна прямая передача сообщений. Если $d^1>d^2$, то формируемая сеть уже относится к классу DTN

(Delay-and-Disruption Tolerant Network) сетей и требует решения дополнительной задачи – маршрутизации передаваемых сообщений.

В данной работе далее рассматривается более общая ситуация, когда $d^1>d^2$, а для решения задачи маршрутизации сообщений используется конкретная специфика формируемой сети. В данном случае структура сети в каждый момент времени представляется состоящей из двух групп узлов. Первую группу узлов образуют АНПА, прикрепленные ко дну и находящиеся, тем самым, в неподвижном состоянии, а вторую группу узлов образуют все остальные АНПА, которые в текущий момент времени находятся в движении. При такой структуризации сети решение задачи маршрутизации сообщений можно рассматривать в зависимости от ситуации:

1 – передача сообщения между узлами, принадлежащими первой группе,

2 – передача сообщения из узла второй группы узлу первой группы, и

3 – передача сообщения из узла первой группы узлу второй группы.

В случае 1 может использоваться следующая возможность. Если агент узла отправителя сообщения знает текущие координаты всех остальных узлов первой группы, то он может построить граф сети узлов этой группы и обычным образом рассчитать маршрут передачи сообщения. При этом, решение задачи маршрутизации в случае 2 сводится к подходу, используемому в случае 1. Более детально эта ситуация рассматривается далее при детальном описании протоколов взаимодействия. Из этого же описания также будет следовать, что ситуации типа 3 в предлагаемой схеме информационного взаимодействия не возникают.

Второй фактор. В работе так же полагается, что устанавливаемые на АНПА модемы могут иметь встроенные цезиевые часы [14]. В этом случае гидроакустические сигналы наряду с передачей информации могут использоваться для измерения расстояний между АНПА и определения текущих координат на основе разностно-дальномерного метода [22].

Третий фактор. Возможности информационного взаимодействия между АНПА и излучателем определяются следующими возможностями и ограничениями. Излучатель может знать (видеть) текущие координаты АНПА без какого-либо участия последних, а также посылать сообщения АНПА. Однако передача сообщений от АНПА к излучателю в общем случае является невозможной (она является возможной только в частных случаях, когда расстояние между АНПА и излучателем меньше значения параметра d^2). При наличии таких ограничений возникает ряд факторов, из-за которых использование централизованного подхода к управлению АНПА (излучатель управляет группировкой АНПА) представляется невозможным.

3. Схема информационного взаимодействия

В рассматриваемой схеме информационного взаимодействия предполагается агент одного АНПА, играющего роль текущего лидера группы. Им определяется агент одного из АНПА, находящихся в состоянии *f*, т.е. прикрепленных ко дну. В качестве возможного варианта по предварительному соглашению лидером может являться агент АНПА, находящегося в текущий момент времени в средней позиции передового ряда. Под передовым понимается ряд АНПА, который последний по времени перешел в состояние *f*. Соответственно при смене передового ряда очевидным образом происходит смена текущего лидера группы.

Далее в статье полагается, что в каждый момент времени имеется один текущий лидер. Но для обеспечения надежности у лидера может рассматриваться дублер, который выполняет те же самые функции, что и лидер.

*Протокол 1***.** Схема протокола приведена на рисунке 2.

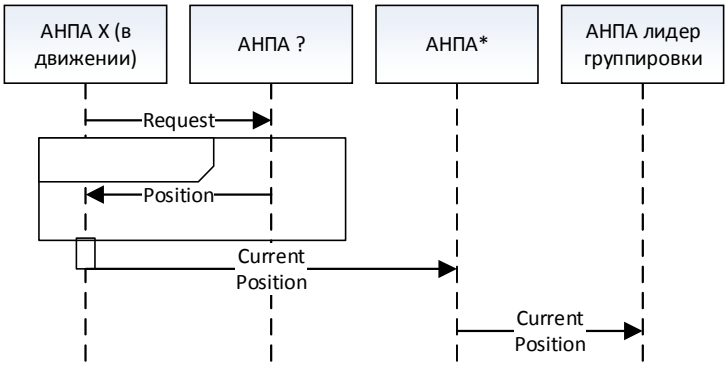


Рис. 2. Протокол 1

Первой целью данного протокола является определение текущих координат АНПА, находящихся в движении. Инициаторами протоколов данного типа являются агенты движущихся АНПА. Для инициирования протокола агент АНПА X посылает в широковещательном режиме сообщение *Request.* В сообщении указывается только момент времени его отправки. Подмножество АНПА, которое получает это сообщение, зависит от их текущего положения относительно АНПА X и дальности распространения сигнала.

При получении сообщения *Request* агент реагирует на него в зависимости от текущего состояния своего АНПА. Если АНПА также находится в движении, то, определяя время получения сигнала, агент определяет текущее расстояние до АНПА X, оценивает возможность столкновения, и, если такая возможность существует, то инициирует координацию дальнейшего движения своего АНПА и АНПА X. Протокол взаимодействия агентов этих АНПА в данной статье в виду ограничения на объем не рассматривается.

Если АНПА находится в состоянии *f,* т.е. прикреплен ко дну, агент посылает ответное сообщение *Position,* в котором указывает следующий набор данных:

– время отправки сообщения,
– координаты нахождения АНПА,
– выполнялось ли или еще нет уточнение координат по протоколу 3, и
– маршрут доставки сообщений до агента, текущего лидера группировки.

Необходимо заметить, что в соответствии с предлагаемой схемой информационного взаимодействия каждый АНПА на основе разностно-дальномерного метода по мере необходимости периодически определяет и/или уточняет свои текущие координаты. Когда АНПА находится в движении, исходными данными являются координаты двух АНПА, прикрепленных ко дну. После прикрепления ко дну исходными данными являются координаты излучателя в два различных момента времени, которые становятся известными агенту АНПА X после получения соответствующих сообщений по протоколу 3. *Протокол 3* сводится к тому, что излучатель с заданной периодичностью посылает в широковещательном режиме агентам всех АНПА свои текущие координаты и момент времени отправки данного сообщения.

Второй способ обеспечивает более высокую точность определения координат, но реализуется только спустя некоторое время после момента времени прикрепления АНПА ко дну. Это время определяется получением двух сообщений с координатами излучателя в соответствующие моменты времени.

После отправки сообщения *Request* агент АНПА X получает несколько ответных сообщений *Position.* В связи с предыдущим пояснением для определения текущих координат в первую очередь выбираются сообщения от агентов АНПА, которые уже успели уточнить свои текущие координаты по протоколу 3.

Далее агент АНПА X пересылает свои текущие координаты (сообщение *Current Position*) агенту, текущему лидеру группы. Для пересылки данных координат в качестве ретранслятора выбирается агент АНПА с наиболее коротким маршрутом доставки сообщения текущему лидеру группировки. При получении сообщения *Position* агент выбранного АНПА инициирует пересылку этого сообщения по соответствующему маршруту. При этом для обеспечения надежности его доставки в схеме протокола на рисунке 1 может быть введен дополнительный фраг-

мент протокола, который реализует следующую логику. При пересылке сообщения на каждом отрезке маршрута должно выполняться ответное подтверждение о получении сообщения. Если в течение определенного времени подтверждение не получено, делается вывод о возможном выходе соответствующего АНПА из строя, и агент соответствующего АНПА, который не получил подтверждения, делает пересчет маршрута доставки сообщения опять-таки в рамках опорного фрагмента сети связи, формируемого неподвижными АНПА, прикрепленными ко дну.

На основе приведенного описания протокола следует привести дополнительные пояснения к рисунку 1. В рамках различных протоколов 1 агент АНПА рассматривается в различных ролях. Когда АНПА находится в состоянии *a, b* или *f,* т.е. в состоянии движения, агент этого АНПА является инициатором протокола. При этом он может инициировать этот протокол неоднократно, по мере необходимости уточнения своих текущих координат. Когда АНПА находится в состоянии *d,* т.е. прикреплен ко дну, то агент этого АНПА является респондентом протокола: пересылает в ответном сообщении *Position* свои текущие данные, может быть ретранслятором передачи сообщения *Current Position* или получателем этого сообщения, если он является текущим лидером группировки.

При выполнении протоколов 1 для сокращения трафика передаваемых сообщений может использоваться возможность, которая заключается в следующем. Агент движущегося АНПА может принимать и использовать в своих целях сообщения *Position*, которые посылаются в ответ на сообщения *Request,* посылаемые агентами других АНПА.

Протокол 2.

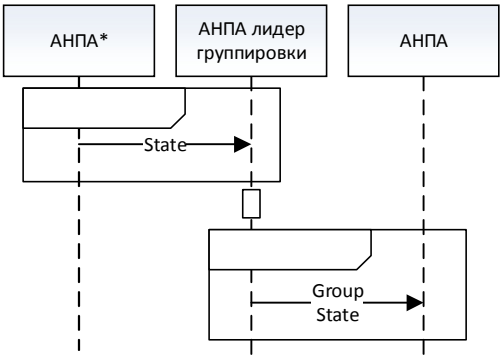


Рис. 3. Протокол 2

После перехода в состояние *c*, т.е. прикрепление ко дну в очередной маршрутной точке плана, агент каждого АНПА* посылает текущему лидеру группировки сообщение *State*, в котором указывает место АНПА в ряду.

В силу выполнения протоколов 1 текущий лидер группировки имеет актуальную информацию о состоянии всех движущихся АНПА. На основании этого он, в частности, знает, какие АНПА должны прикрепиться ко дну, и послать сообщение *State*, а какие АНПА вышли из строя. При получении такого сообщения от последнего АНПА, прибывшего в очередную маршрутную точку плана, текущий лидер группы идентифицирует окончание передвижения ряда. При этом у лидера, и пока только у него, формируется набор данных, описывающих текущее состояние группы. Этот набор данных, в частности, позволяет идентифицировать не занятые маршрутные точки плана и вышедшие из строя АНПА. В соответствии с этим следующим шагом лидера группы является доведение этих данных до всех остальных АНПА группы. Для этого лидер группы посылает агентам всех АНПА сообщение *Group State,* содержащее этот набор данных. Для сокращения трафика передаваемых сообщений в рамках этой рассылки агент-лидер в соответствующих сообщениях указывает необходимость ретрансляции этого же сообщения агентам других АНПА, которые в данный момент находятся за пределами распространения гидроакустического сигнала.

Получение сообщения *Group State* агентами АНПА является сигналом для автономного определения нового текущего лидера группы. Им, в соответствии с предопределенным правилом, становится агент АНПА, занимающего среднее место в ряду, который перешел в состояние *c.* Для агентов АНПА последнего ряда это же сообщение также является сигналом для начала передвижения в последующие маршрутные точки плана.

4. Схемы построения, передвижения и перестроения группы АНПА

В данном разделе рассматриваются различные схемы построения, передвижения и перестроения группы АНПА, и приводятся аналитические оценки влияния тех или иных схем на качество и время выполнения миссии сейсморазведки заданного района.

Под схемой построения группы АНПА главным образом подразумевается задание значений параметров, определяющих форму строя: $n*m$, где n количество рядов, а m количество шеренг в строю. Также в качестве исследуемых параметров необходимо рассматривать расстояние d^0 между соседними АНПА в рядах и шеренгах, так как на его основе рассчитываются парные расстояния между АНПА, и с учетом дальности распространения гидроакустических сигналов d^2 определяется маршрутизация передачи сообщений по сети АНПА.

Под схемой передвижения подразумеваются два фактора. Под первым фактором подразумевается количество рядов, одновременно находящихся в движении: один ряд, два или более. Под вторым фактором – схема выполнения поворота группы. Поворот может осуществляться передвижением ряда (рядов) или шеренги (шеренг).

Перестроение группы рассматривается в случаях выхода каких-то АНПА из строя и подразумевает выбор агентами работоспособных АНПА последующих маршрутных точек для перехода в соответствии с той или иной схемой перестроения в зависимости от ситуации.

Различные схемы построения, передвижения и перестроения группы АНПА влекут различные показатели времени и качества выполнения миссии сейсморазведки в целом.

Время выполнения миссии предопределяется тремя временными параметрами:

– t^1 – время выполнения сеанса сейсморазведки,

– t^2 – время выполнения передвижения ряда АНПА в последующие маршрутные точки плана в случае линейного движения, и

– t^3 - время выполнения передвижения ряда АНПА в последующие маршрутные точки плана в случае поворота группы.

Все три параметра зависят от значений параметров n, m, d^0. При этом время t^1 также зависит от скорости передвижения излучателя, а времена t^2 и t^3 - от скорости передвижения АНПА, а также от схемы передвижения (схемы выполнения поворота) в случаях поворота группы, и от схемы перестроения группы в случае появления выбывших из строя АНПА. Исходя из этого, можно оценить минимальное время выполнения миссии T^M при условии, что все АНПА на протяжении всей миссии остаются в рабочем состоянии. Это можно сделать с помощью формулы

$$T^M = k^L*min(t^1,t^2) + k^T*min(t^1,t^3),$$

где k^L – количество сеансов сейсморазведки при линейном передвижении группы, а k^T – количество сеансов сейсморазведки при поворотах группы, предполагаемые планом выполнения миссии. Если появляются АНПА, вышедшие из строя, то время выполнения миссии T^M как минимум увеличивается в зависимости от выбираемых схем перестроения группы.

Качество выполнения миссии главным образом определяется эффективностью сейсморазведки, показателем которой может служить всеазимутальность направлений прихода отраженных сигналов, принятых АНПА в маршрутных точках плана. При этом для оценки качества выполнения миссии также можно рассматривать дополнительный показатель, который оказывает непосредственное влияние на значение основного показателя в маршрутных точках плана: количество сеансов сейсморазведки, в течение которых в маршрутной точке принимались сигналы, т.е. в точке находился АНПА. В связи с этим целесообразно обратить внимание на два фактора.

Фактор 1. При повороте группы даже в случае, когда все АНПА находятся в рабочем состоянии, имеются маршрутные точки, в которых сбор данных выполняется в ограниченном количестве сеансов сейсморазведки.

Фактор 2. При определенных схемах перестроения группы, в случае появления выбывших из строя АНПА, могут быть маршрутные точки, в которых вообще не производится сбор данных.

Для пояснения фактора 1 можно использовать пример, представленный на рисунке 4. В данном случае поворот группы выполняется передвижением в последующие точки последнего ряда. В маршрутных точках указано количество сеансов сейсморазведки, в течение которых в них проводится сбор данных. При данном построении группы, $n=4$ и $m=3$, как правило, во всех

точках сбор данных выполняется в течение трех сеансов сейсморазведки, но в случае поворота есть три точки, в которых сбор данных выполняется в течение двух или одного сеанса.

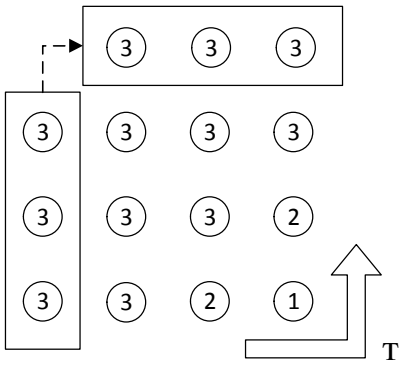


Рис.4. Качество сбора данных в маршрутных очках при повороте группы

Очевидно, что наименьшее время и наилучшее качество выполнения миссии обеспечиваются в том случае, когда все АНПА остаются в рабочем состоянии. При этом можно привести модель выполнения миссии, которая также может обеспечивать наилучшее время и качество выполнения миссии даже в случае выхода из строя определенного количества АНПА. В данной модели полагается, что $n=m+2$. В рамках каждого сеанса сейсморазведки собираются данные на множестве точек $m*m$. Во время сеанса сейсморазведки в последующие маршрутные точки плана происходит передвижение одного ряда АНПА, последнего. Второй, ряд АНПА, предпоследний, рассматривается как резервный, и АНПА этого ряда во время очередного сеанса сейсморазведки остаются прикрепленными ко дну, но не выполняют сбор данных, и выполняют переход в последующие маршрутные точки плана в течение последующего сеанса. При выходе каких-то АНПА из строя их заменяют АНПА резервного ряда в соответствии с достаточно очевидной схемой перестроения, в основе которой рассматривается следующий подход. Выбывающий из строя АНПА заменяет АНПА резервного ряда из той же шеренги. Если такой АНПА уже использован в замене выбывшего из строя АНПА, то выбирается АНПА из другой шеренги, которая определяется на основе априори определенного правила. Такая модель может обеспечивать достижение наилучшего времени и качества выполнения миссии, но за счет использования резервных АНПА и до тех пор, пока количество выбывших из строя АНПА не превышает количество резервных АНПА.

5. Анализ схемы информационного взаимодействия

Основным объектом исследований в данной статье является схема информационного взаимодействия в группе АНПА. Разработка и выбор наиболее рациональных схем построения, передвижения и перестроения группы АНПА является объектом исследований последующего этапа работы. В связи с этим в данном разделе рассматриваются некоторые свойства предложенной схемы информационного взаимодействия, характеризующие ее возможную взаимосвязь с различными вариантами построения, передвижения и перестроения группы АНПА.

Основной целью предложенной схемы информационного взаимодействия является обеспечение агентов всех АНПА актуальными данными, описывающими текущее состояние группы в целом. Содержание этих данных является необходимым и достаточным для того, чтобы агент каждого АНПА в соответствии с предопределенными правилами мог автономно планировать дальнейшие действия своего АНПА в зависимости от текущей ситуации.

Эти данные формирует текущий лидер группы после прикрепления ко дну в очередных маршрутных точках АНПА передвигающегося ряда. Далее лидер группы пересылает эти данные агентам всех остальных АНПА. При этом агенты АНПА, которые закончили участие в очередном сеансе сейсморазведки и находятся в состоянии *c* (рис.1), используют эти данные для автономного выбора последующей маршрутной точки и определения схемы передвижения в эти точки. После принятия такого решения АНПА выполняют переход в эти маршрутные точки. Таким образом, информация о выходе из строя одного или нескольких их аппаратов ста-

новится доступной агентам всех АНПА в определенные моменты времени, и реагируют на эту информацию только агенты тех АНПА, которым надо определить цель предстоящего передвижения. Такой подход определяет определенный уровень оперативности реагирования группы АНПА на выход из строя каких-то аппаратов.

Однако предложенная схема информационного взаимодействия содержит потенциал для повышения уровня оперативности реагирования на такие события. Например, агенты передвигающихся АНПА могут получать информацию о выходе каких-то из них из строя более оперативно, т.е. за какое-то время до заглубления и прикрепления ко дну. Эта возможность может быть реализована за счет незначительных доработок протокола 1 (рис.2).

Заключение

В статье приводится описание схемы информационного взаимодействия группы АНПА при выполнении миссии сейсморазведки. Данная схема предполагает и обеспечивает автономное планирование поведения аппаратов в группе. Она может использоваться при различных схемах построения и передвижения группы АНПА, и главным образом – при использовании различных схем перестроения группы в случае выбывания из строя каких-то АНПА.

На последующем этапе исследований данная схема будет использоваться для разработки, моделирования и выбора схем построения, передвижения и перестроения группы АНПА, обеспечивающих наилучшие показатели времени и качества выполнения миссии сейсморазведки в целом.

Работа проводилась при поддержке гранта РФФИ 17-08-00666

ЛИТЕРАТУРА

1. Каляев И.А., Капустян С.Г. Проблемы группового управления роботами // Мехатроника, автоматизация, управление. – №6. – 2009. – С. 33-40.
2. Юревич Е.И. О проблеме группового управления роботами // Мехатроника, автоматизация, управление. – 2004. – № 2. – С. 9-13.
3. Каляев И.А. Модели и алгоритмы коллективного управления в группах роботов / И.А. Каляев, А.Р. Гайдук, С.Г. Капустян. – М.: Физматлит, 2009. – 280 с.
4. Куржанский А.Б. О задаче группового управления в условиях препятствий // Тр. Института математики и механики РАН. - 2014. -Т.20. - №3. -С.166-179.
5. Городецкий В.И., Грушинский М.С., Хабалов А.В. Многоагентные системы (обзор) – Новости искусственного интеллекта, 1998, №2, с.64-116.
6. Будаев Д.С., Вощук Г.Ю., Гусев Н.А., Майоров И.В., Мочалкин А.Н. Разработка системы согласованного управления группой беспилотных аппаратов с применением мультиагентных технологий // «Известия ЮФУ. Технические науки», № 10(171), 2015. – С. 18-28.
7. Городецкий В. И., Карсаев О. В., Самойлов В. В., Серебряков С. В. Прикладные многоагентные системы группового управления // Искусственный интеллект и принятие решений. 2009. № 2. С. 3–24.
8. Al-Khatib H., Antonelli G., Caffaz A., Caiti A., Casalino G., de Jong I.B., Duarte H., Indiveri G., Jesus S., Kebkal K., Pascoal A., Polani D. The widely scalable mobile underwater sonar technology (WiMUST) project: an overview // Proceedings of MTS/IEEE Oceans ’15, Genova, Italy, May 18-21 2015, pp. 1–5. [Online]. Available: http://dx.doi.org/10.1109/OCEANS-Genova.2015.7271688.
9. Klyubin, A.S., Polani, D., & Nehaniv, C.L. Empowerment: A universal agentcentric measure of control. In: IEEE Congress on Evolutionary Computation (Volume:1), 2005. pp. 128-35. Edinburgh, Scotland: IEEE. http://dx.doi.org/10.1109/cec.2005.1554676.
10. Klyubin, A.S., Polani, D., & Nehaniv, C.L. Keep your options open: An informationbased driving principle for sensorimotor systems. 2008. PloS ONE. 3(12):e4018. http://dx. doi.org/10.1371/journal.pone.0004018.
11. Ghabcheloo, R., Aguiar, A.P., Pascoal, A.M., Silvestre, C., Kaminer, I., & Hespanha, J. 2009. Coordinated Path-Following in the Presence of Communication Losses and Time Delays. SIAM J Control Optim. 48(1):234- 65. http://dx.doi.org/10.1137/060678993.
12. De Palma, D., Indiveri, G., & Parlangeli, G. 2015a. Multi-vehicle relative localization based on single range measurements. In: 3rd IFAC Workshop on Multivehicle Systems (MVS 2015). Genoa, Italy: IFACPapersOnLine. 48(5):17-22. http://dx.doi. org/10.1016/j.ifacol.2015.06.457.
13. De Palma, D., Indiveri, G., & Pascoal, A.M. 2015b. A Null-Space-Based Behavioral Approach to Single Range Underwater Positioning. In: 10th IFAC Conference on Manoeuvring and Control of Marine Craft (MCMC 2015). Copenhagen, Norway: IFAC-PapersOnLine. 48(16):55-60.
14. Кебкал К.Г., Кебкал А.Г, Глушко Е.В., Кебкал В.К., Себастио Л., Паскуаль А., Рибейро Дж., Сильва Г., Рибейро М., Индивери Дж. Гидроакустические модемы с интегрированными цезиевыми часами для задач подводного

позиционирования автономных необитаемых подводных аппаратоВ // International Conference on Marine Robotics in Ocean Exploration, MarineRobotics2017, October 9-11, Saint-Petersburg, Russia. c.1-10.
15. Мартынова Л.А. Метод координации функционирования агентов в мультиагентной системе управления АНПА // Труды Международной конференции по судостроению и океанотехнике: – СПб: СПбГМТУ, 2016, с. 470-480.
16. Мартынова Л.А. Метод согласованного поведения излучателя и автономных необитаемых подводных аппаратов для эффективного ведения сейсморазведки // Информационно-управляющие системы. 2017. № 1(86). С. 83–92.
17. Мартынова Л.А. Инструментарий для исследований эффективности ведения сейсморазведки с использованием автономных необитаемых подводных аппаратов // Информационно-управляющие системы. 2017. №2 (87). С. 77-87.
18. Мартынова Л.А., Розенгауз М.Б. Определение эффективного поведения группы АНПА в сетецентрической системе освещения подводной обстановки // Информационно-управляющие системы. 2017. №3 (88). С. 47-57.
19. Мартынова Л.А. Дифференциальный метод позиционирования автономного необитаемого подводного аппарата при ведении сейсморазведки // Информационно-управляющие системы. 2017. №4 (89). С. 77-85.
20. Мартынова Л.А, Конюхов Г.В., Пашкевич И.В., Рухлов Н.Н. Технология разработки моделей и алгоритмов системы управления АНПА при выполнении сложной миссии // Материалы двенадцатой Всероссийской научно-практической конференции "Перспективные системы и задачи управления" 3-7 апреля 2017г. в п.Домбай Карачаево-Черкесской республики. – Ростов-на-Дону. – 2017. – С.117-128.
21. Кожемякин И.В., Рыжов В.А., Семенов Н.Н., Чемоданов М.Н. Исследование влияния скорости и дальности связи между элементами мультиагентной сети подводных роботов на ее устойчивость к изменению условий работы // Известия ЮФУ. Технические науки. 2018. №1. С. 227-241.
22. Черняк В.С. Многопозиционная радиолокация. – М.: Радио и связь, 1993. – 416 с.

L.A. Martynova (Concern CSRI Elektropribor, Saint Petersburg), **O.V. Karsaev** (SPIIRAS, Saint Petersburg)

Group control of autonomous unmanned underwater vehicles' multi-agent system in the conduct of all-weather seismic survey

The article describes a model of autonomous command behavior of the AUV group performing the mission of seismic exploration of a given sea bed area. The model of behavior is developed on the basis of multi-agent approach. The main object of research in the article is the scheme of information interaction. The purpose of information exchange is to provide all AUV with operational information about the current state of the group. At that, AUV agents can use this information for autonomous planning of the AUV subsequent actions. Autonomous planning of actions is performed in accordance with a priori defined schemes of construction, movement and reconstruction of the AUV group topology. Schema reconstruction considered in cases of failure of any of AUV.

ПОТОКОВЫЕ РЕКУРРЕНТНЫЕ НЕЙРОННЫЕ СЕТИ С УПРАВЛЯЕМЫМИ ЭЛЕМЕНТАМИ

В. Ю. Осипов[1], **В. В. Никифоров**[2]
(*Санкт-Петербургский институт информатики и автоматизации РАН (СПИИРАН)
199178, С.-Петербург, 14 линия, 39, тел. (812)328-08-87, nik@iias.spb.su*)

Аннотация

Исследуются особенности линейных и многовитковых потоковых рекуррентных нейронных сетей с управляемыми элементами. Уточняются правила ассоциативной обработки такими сетями сигналов в виде совокупностей последовательностей единичных импульсов (образов). Анализируются вклады различных групп синапсов в такую обработку. Приводятся примеры по одновременному ассоциативному восстановлению недостающих и подавлению ложных элементов в обрабатываемых сигналах. На основе полученных результатов моделирования обоснованы пути снижения структурной сложности этих нейронных сетей и расширения возможностей их ассоциативной памяти.

Ключевые слова: структура нейронных сетей, синапсы, сигналы, ассоциативная память.

Введение

Одним из перспективных направлений в области интеллектуальной обработки разнородной информации выступает развитие и применение искусственных рекуррентных нейронных сетей (РНС). Желательно, чтобы РНС позволяли не только быстро, но и глубоко обрабатывать поступающие в них потоки сигналов. При этом они должны быть относительно простыми и прозрачными в структурном и функциональном плане. Традиционные нейросетевые решения во многом не удовлетворяют этим требованиям. Известные РНС реального времени на базе персептрона обеспечивают быструю, но не глубокую обработку входных сигналов. К таким сетям относятся рекуррентный многослойный персептрон, сеть Эльмана, рекуррентная сеть реального времени и другие. В противовес им известные самоорганизующиеся рекуррентные сети Хопфилда, Коско, Кохенена и другие ориентированы на глубокую обработку сигналов, но не удовлетворяют требованиям реального времени [1, 2]. Кроме этого эти РНС во многом не прозрачны в структурном плане и несовершенны в функциональном отношении.

Расширенными возможностями по ассоциативной обработке потоков сигналов обладают развиваемые авторами рекуррентные нейронные сети с управляемыми элементами [3]. За счет реализации в них в зависимости от текущих состояний слоев пространственных сдвигов сигналов, передаваемых от слоя к слою, удается наделить эти сети прозрачными линейными, спиральными и другими видами логических структур. За счет управления параметрами пространственной расходимости сигналов в таких РНС осуществима, как быстрая, так и глубокая ассоциативная обработка информации [4].

Перед подачей сигналов в эту РНС они подлежат прямому преобразованию. Согласно ему сначала сигналы могут раскладываться, в общем случае на пространственно-частотные составляющие. Затем каждая составляющая преобразуется в последовательность единичных импульсов (образов) с частотой и фазой следования, как функциями от амплитуды и фазы составляющей. В результате, если рассматривать процесс в дискретном времени, то на вход РНС подаются последовательности совокупностей единичных образов (импульсов).

Такие совокупности единичных образов не только передаются от слоя к слою, но и продвигаются вдоль слоев в заданных направлениях за счет реализуемых пространственных сдвигов. При продвижении по сети сигналы ассоциируются друг с другом и с ранее запомненной информацией с учетом управления направленностью ассоциаций в зависимости от текущих состояний слоев. При этом сохраняется возможность обратного преобразования последовательных совокупностей единичных образов на выходе РНС в соответствующие им исходные сигналы. Такие РНС в силу специфики их функционирования можно также называть потоковыми РНС.

[1] Доктор технических наук, руководитель лаборатории.
[2] Доктор технических наук, главный научный сотрудник.

Несмотря на ряд полученных результатов, многие свойства потоковых РНС с управляемыми элементами во многом остаются не исследованными.

В докладе уточняются структурные особенности таких РНС, анализируются вклады различных групп синапсов в ассоциативную обработку сигналов. Предлагаются пути снижения структурной сложности этих РНС, способы одновременного восстановления в РНС недостающих и подавления ложных элементов в обрабатываемых сигналах.

Архитектура потоковых рекуррентных нейронных сетей

Обобщенную структуру потоковой рекуррентной нейронной сети (РНС) с управляемыми элементами можно представить в виде рис. 1.

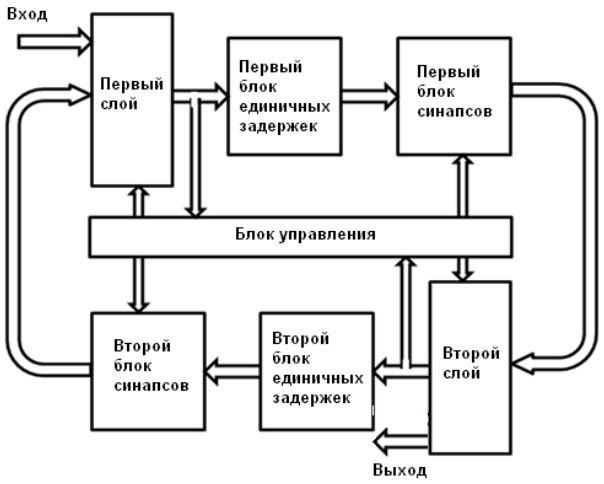


Рис. 1. Обобщенная структура рекуррентной нейронной сети с управляемыми элементами

Специфическая особенность такой РНС в том, что она может наделяться различными логическими структурами и обеспечивать управляемую ассоциативную обработку потоков сигналов.

Слои нейронов. Рассматриваемые потоковые РНС состоят из нейронов $N_{i,j,k}$, размещаемых в вершинах трехмерной решетки (рис. 2) с целочисленными координатами:

$$0 \le i < 2, \quad 0 \le j < J, \quad 0 \le k < K, \qquad (1)$$

где J – длина слоя нейронов, K – ширина слоя. Набор значений трех индексов i, j, k, отвечающих неравенствам (1), однозначно указывают имеющийся в сети конкретный нейрон. Подмножество $L_0 = \{N_{0,j,k}\}$ всех нейронов, со значением $i = 0$ составляет *верхний слой* нейронов. Подмножество $L_1 = \{N_{1,j,k}\}$ всех нейронов, со значением $i = 1$ составляет *нижний слой* нейронов.

В конкретный момент времени t текущее состояние нейрона $N_{i,j,k}$ характеризуется параметром $S_t(N_{i,j,k})$ — *кодом состояния*, принимающим одно из трех значений: 1 – активное; 0 – пассивное; -1 – невосприимчивое состояние нейрона.

Поля, составляющие слой нейронов. Каждый из двух слоев сети строится в виде последовательности идентичных полей $F_{l,x}$ (l — номер слоя, которому принадлежит поле; x — порядковый индекс поля в его слое). Так, слой L_0 состоит из полей $F_{0,0}$, $F_{0,1}$, …, $F_{0,((J/fl)-1)}$, слой L_1 — из полей $F_{1,0}$, $F_{1,1}$, …, $F_{1,((J/fl)-1)}$ (значение fl указывает длину каждого из полей сети, отношение J / fl соответствует числу полей в каждом из слоев L_0 и L_1). Поле $F_{l,x}$ содержит нейроны $N_{l,j,k}$ со значениями индекса j, отвечающими неравенствам $x \cdot fl \le j < x \cdot (fl+1)$.

Входной и выходной модули. В левом верхнем углу рис. 2 изображен входной модуль $E(t)$. Структура модуля в точности повторяет структуру каждого из полей. Модуль $E(t)$ содержит $fl \cdot K$ элементов:

$$E(t) = \begin{vmatrix} e_{0,0}(t), & e_{0,1}(t), & \dots & , e_{0,fl-1}(t) \\ e_{1,0}(t), & e_{1,1}(t), & \dots & , e_{1,fl-1}(t) \\ \dots & \dots & \dots & \dots \\ e_{K-1,0}(t), & e_{K-1,1}(t), & \dots & , e_{K-1,fl-1}(t) \end{vmatrix}.$$

Элементы $e_{j,k}(t)$ принимают значения 0 или 1. Значение элемента $e_{j,k}(t)$ отражает текущее состояние соответствующего элемента внешнего оборудования. Входной модуль $E(t)$ не является объектом того же типа, что и поля $F_{l,x}$ (несмотря на совпадение их структур) — поле содержит нейроны $N_{i,j,k}$, а входной модуль $E(t)$ содержит элементы $e_{j,k}(t)$, не являющиеся нейронами. Нейрон может иметь несколько входных (и выходных) синаптических связей с другими нейронами, а элемент $e_{j,k}(t)$, имеет только возбуждающую связь с нейроном $N_{0,j,k}$ (пример такой возбуждающей связи от $e_{0,2}$ к $N_{0,0,2}$ приведен на рис. 1). Для акцентирования различия между входным модулем и полями слоев РНС на рис. 2 входной модуль ограничен жирной рамкой.

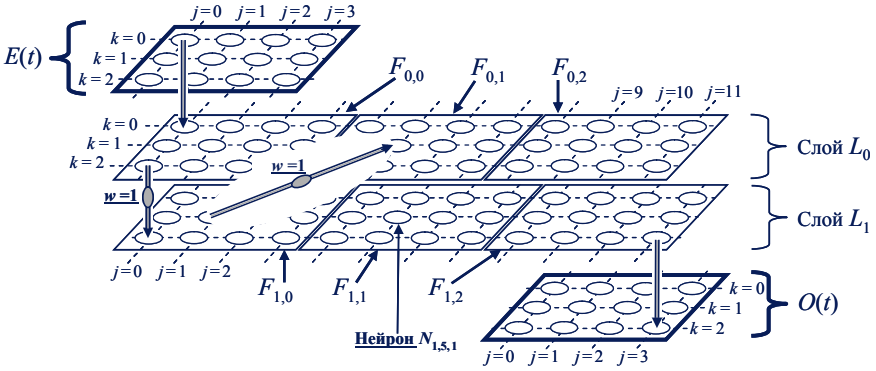


Рис. 2. Потоковая РНС с параметрами J=12, K=3, fl = 4

В правом нижнем углу рис. 1 изображен выходной модуль $O(t)$. Структура модуля в точности повторяет структуру каждого из полей сети. Модуль $O(t)$ содержит $fl \cdot K$ элементов:

$$O(t) = \begin{vmatrix} o_{0,0}(t), & o_{0,1}(t), & \dots & , o_{0,fl-1}(t) \\ o_{1,0}(t), & o_{1,1}(t), & \dots & , o_{1,fl-1}(t) \\ \dots & \dots & \dots & \dots \\ o_{K-1,0}(t), & o_{K-1,1}(t), & \dots & , o_{K-1,fl-1}(t) \end{vmatrix}.$$

Элементы $o_{j,k}(t)$ принимают значения 0 или 1. Значение $o_{j,k}(t)$ отражает уровень управляющего сигнала, направляемого в момент t на соответствующий элемент внешнего оборудования. Так же, как и входной модуль $E(t)$, выходной модуль $O(t)$ не является объектом того же типа, как поля, составляющие слои сети.

Элемент $o_{j,k}(t)$ имеет с сетью только одну сигнальную связь, идущую от нейрона $N_{1,(J-fl+j),k)}$ (пример такой связи от нейрона $N_{1,11,0)}$ к элементу $o_{3,0}$ приведен на рис. 2). Для акцентирования различия между выходным модулем и полями слоев нейронной сети на рис. 2 выходной модуль ограничен жирной рамкой.

Синаптические связи. При $i \neq i'$ нейроны $N_{i,j,k}$ и $N_{i',j',k'}$ могут быть связаны направленной синаптической связью ($N_{i,j,k}$, $N_{i',j',k'}$), обеспечивающей передачу возбуждающих импульсов от $N_{i,j,k}$ к $N_{i',j',k'}$. Для синаптической связи ($N_{i,j,k}$, $N_{i',j',k'}$) нейрон $N_{i,j,k}$ является *передающим нейроном*, нейрон $N_{i',j',k'}$ – *принимающим нейроном*. Принимающий нейрон может возбуждаться путем передачи через синаптическую связь возбуждающего импульса от передающего нейрона. Синаптическая связь ($N_{i,j,k}$, $N_{i',j',k'}$) является *входной связью* для нейрона $N_{i',j',k'}$, для $N_{i,j,k}$ - *выходной* связью. Накладываемое на синаптическую связь ($N_{i,j,k}$, $N_{i',j',k'}$) ограничение $i \neq i'$ означает отсутствие синаптических связей между нейронами, принадлежащими одному и тому же слою. Для любой имеющейся в потоковой РНС синаптической связи передающий и принимающий нейроны принадлежат разным слоям.

Текущее (в момент времени t) состояние синаптической связи ($N_{i,j,k}$, $N_{i',j',k'}$) характеризуется значением параметра $w_t(N_{i,j,k}, N_{i',j',k'})$ – *весом* синаптической связи. Значение $w_t(N_{i,j,k}, N_{i',j',k'})$ может изменяться от 0 до 1.

В потоковых РНС различаются два вида синаптических связей: *запоминающие* и *направляющие* синптические связи.

Направляющие синаптические связи задают путь движения потока входных данных в РНС. Каждый нейрон $N_{0,j,k}$ верхнего слоя имеет одну исходящую *вертикальную* направляющую связь ($N_{0,j,k}$, $N_{1,j,k}$). Связь ($N_{0,0,0}$, $N_{1,0,0}$) на рис. 2 является примером вертикальной направляющей связи. Каждый нейрон $N_{1,j,k}$ нижнего слоя имеет одну исходящую *диагональную* направляющую связь ($N_{1,j,k}$, $N_{0,j+fl,k}$). Связь ($N_{1,1,1}$, $N_{0,5,1}$) на рис. 2 является примером *диагональной* направляющей связи. Вертикальные и диагональные направляющие связи задают маршрут движения обрабатываемых входных объектов (эталонов или их искаженных образцов) от входного модуля сети вдоль слоев ее полей к выходному модулю**.** В режиме обучения эталон, поступающий из входного модуля в поле $F_{0,0}$, продвигается по направляющим связям от одного поля к другому вдоль следующей цепочки полей:

$$F_{0,0} \rightarrow F_{1,0} \rightarrow F_{0,1} \rightarrow F_{1,1} \rightarrow F_{0,2} \rightarrow \ldots \rightarrow F_{0,((J/fl)-1)} \rightarrow F_{1,((J/fl)-1)}.$$

Форма конкретного эталона, поступающего во входной модуль E, определяется указанием подмножества элементов $\{e_{j,k} = 1\}$. Каждому из этих элементов в поле $F_{l,x}$ соответствует нейрон $N_{l,j+x\bullet fl,k}$. Подмножество нейронов поля $F_{l,x}$, соответствующее таким образом подмножеству $\{e_{j,k} = 1\}$, назовем *следом* этого конкретного эталона в поле $F_{l,x}$.

Пачки направляющих связей. Вертикальные направляющие связи ($N_{0,j,k}$, $N_{1,j,k}$) и ($N_{0,j',k'}$, $N_{0,j',k'}$) являются *соседними* направляющими, если нейроны $N_{0,j,k}$ и $N_{0,j',k'}$ принадлежат одному и тому же полю. Аналогично диагональные направляющие связи ($N_{1,j,k}$, $N_{0,j+fl,k}$) и ($N_{1,j',k'}$, $N_{0,j'+fl,k'}$) являются соседними, если нейроны $N_{1,j,k}$ и $N_{1,j',k'}$ принадлежат одному и тому же полю.

Совокупность соседних направляющих связей, являющихся выходными для нейронов из следа конкретного эталона в поле $F_{l,x}$ обеспечивает выполнение очередного шага продвижения этого эталона от поля $F_{l,x}$ к следующему полю представленной выше цепочки. Совокупность таких соседних направляющих назовем *пачкой направляющих* по данному конкретному эталону. Это структурное понятие удобно использовать при анализе движения обрабатываемых образцов вдоль потоковой РНС. Пачка направляющих соединяет след эталона в поле $F_{l,x}$ со следом того же эталона в следующем (по цепочке) поле.

Все синаптические связи, не относящиеся к разряду направляющих, являются *запоминающими синаптическими связями*.

Правила работы потоковой РНС

Потоковые РНС работают в дискретном режиме: изменение значений их динамических параметров (*состояний* $S_t(N_{i,j,k})$ нейронов и *весов* $w_t(N_{i,j,k}, N_{i',j',k'})$ запоминающих синапсов) возникают только в моменты времени t_0, t_1, t_2, … (*такты* работы сети). Положим, что такты работы сети следуют равномерно через 1 единицу времени. То есть, символы t_x и t_y могут рассматриваться как целые числа, $(t_x - t_y) = (x - y)$.

Исходные значения параметров сети. Каждый из нейронов инициализируется со значением состояния $S_0(N_{i,j,k}) = 0$ (все нейроны приводятся в пассивное состояние). Запоминающие синапсы инициализируются со значениями веса $w_0(N_{i,j,k}, N_{i',j',k'}) = 0$. Направляющие синаптические связи инициализируются со значениями веса $w_0(N_{i,j,k}, N_{i',j',k'}) = 1$.

Правила изменения состояний нейронов. В потоковых РНС нейроны слоев L_0 и L_1 попеременно играют роль передающих и принимающих нейронов. На четных тактах t_x нейроны слоя L_0 играют роль принимающих нейронов, нейроны слоя L_1 — роль передающих нейронов; на нечетных тактах нейроны слоев L_0 и L_1 меняются ролями.

Состояние $S_t(N_{i,j,k})$ каждого из нейронов может изменяться на любом такте t. По отношению к моменту t_x состояние $S_t(N_{i,j,k})$ является *предыдущим* состоянием, если $(t_x - 1) < t \le t_x$; если $t_x < t \le (t_x + 1)$ то $S_t(N_{i,j,k})$ по отношению к моменту t_x является *последующим* состоянием. Аналогичные термины *предыдущий вес* и *последующий вес* используются в отношении значений весов $w_t(N_{i,j,k}, N_{i',j',k'})$ синапсов.

Последующее по отношению к t_x состояние принимающего нейрона $N_{i',j',k'}$ опеделяется тремя факторами:

— предыдущим состоянием тех нейронов $S_t(N_{i,j,k})N_{i,j,k}$, которые являются передающими для $N_{i',j',k'}$;

— предыдущими значениями весов тех синапсов, которые играют роль входных синапсов для $N_{i',j',k'}$;

— принятого для сети значения константы B — порога возбуждения нейрона.

Принимающий нейрон $N_{i,j,k}$ возбуждается, если суммарное возбуждающее воздействие $W_t(N_{i,j,k})$ на его входе будет не меньше порога B возбуждения. Для нейронов $N_{1,j,k}$ нижнего слоя и для нейронов $N_{0,j,k}$ верхнего слоя значения этой функции представляется выражениями:

$$W_t(N_{0,j,k}) = \Sigma_{0 \le j < J} \Sigma_{0 \le k < K} w_t(N_{0,j,k}, N_{1,j',k'})\, s_t(N_{1,j',k'}) + inp(j),$$

$$W_t(N_{1,j,k}) = \Sigma_{0 \le j < J} \Sigma_{0 \le k < K} w_t(N_{1,j,k}, N_{0,j',k'})\, s_t(N_{0,j',k'}),$$

$$s_t(N_{i,j,k}) = \begin{cases} 1 & \text{if } S_t(N_{i,j,k}) = 1, \\ 0 & \text{if } S_t(N_{i,j,k}) < 1, \end{cases} \qquad inp_t(j,k) = \begin{cases} e_{j,k}(t) & \text{if } j < J, \\ 0 & \text{if } j \ge J. \end{cases}$$

Определение весов синапсов. Веса вертикальных $w_0(N_{0,j,k}, N_{1,j,k})$ и диагональных $w_0(N_{1,j,k}, N_{0,j+fl,k})$ синапсов в момент инициализации системы устанавливаются равными 1 и остаются неизменными в ходе работы системы.

Веса $w_t(N_{i,j,k}, N_{i',j',k'})$ запоминающих синапсов инициализируются нулевыми значениями и изменяются в режиме обучения, накапливая информацию о свойствах эталонов. В рамках подхода к построению потоковых РНС, представленного в работе [3], предложено устанавливать значения весов запоминающих синапсов $w_t(N_{i,j,k}, N_{i',j',k'})$ режима обучения в соответствии с выражением:

$$w_t(N_{i,j,k}, N_{i',j',k'}) = \pi_t(N_{i,j,k}, N_{i',j',k'}) \bullet \beta(N_{i,j,k}, N_{i',j',k'}) \bullet \eta(N_{i,j,k}, N_{i',j',k'})$$

В этом выражении $\pi_t(N_{i,j,k}, N_{i',j',k'})$ представляет весовой коэффициент, а $\beta(N_{i,j,k}, N_{i',j',k'})$ и $\eta(N_{i,j,k}, N_{i',j',k'})$ – функции ослабления расходящихся и сходящихся единичных образов (импульсов), соответственно. Величина $\pi_t(N_{i,j,k}, N_{i',j',k'})$ может изменяться 0 до 1. Если синаптическая связь $(N_{i,j,k}, N_{i',j',k'})$ не использовалась для возбуждения нейрона $N_{i',j',k'}$ с участием нейрона $N_{i,j,k}$ на протяжении временного интервала $[0, t)$, то значение $\pi_t(N_{i,j,k}, N_{i',j',k'})$ остается нулевым. Чем чаще эта связь использовалась для возбуждения $N_{i',j',k'}$, тем более значение $\pi_t(N_{i,j,k}, N_{i',j',k'})$ приближается к 1. Значения $\beta(N_{i,j,k}, N_{i',j',k'})$ и $\eta(N_{i,j,k}, N_{i',j',k'})$ зависят от расстояния $D(N_{i,j,k}, N_{i',j',k'})$ между нейронами $N_{i,j,k}$ и $N_{i',j',k'}$. Задавая определенным образом значения $\beta(N_{i,j,k}, N_{i',j',k'})$ можно формировать различные логические структуры РНС. На рис. 3*а*, *б* представлены два варианта структур потоковых РНС. На рис. 3*а* представлен план линейной потоковой РНС (вид сверху на слой L_0, только нейроны верхнего слоя изображены на плане). На осях X и Y отмечены значения координат $X(N_{i,j,k})$ и $Y(N_{i,j,k})$ двумерной решетки, в узлах которой размещены нейроны сети. Для линейного варианта эти координаты задаются выражениями

$$X(N_{i,j,k}) = j, \quad Y(N_{i,j,k}) = k.$$

Пример многовитковой потоковой РНС (конкретно — сети, состоящей из трех витков) представлен на рис. 3*б*. Каждый виток содержит cl полей (и, соответственно, $cl \bullet fl \bullet K$ нейронов). В рамках каждого из витков с четными номерами (coil_0, coil_2, ...) значение абсциссы $X(N_{i,j,k})$ нейрона $N_{i,j,k}$ растет с ростом значения индекса j в соответствии с выражением

$$X(N_{i,j,k}) = \mathrm{mod}_{cl \bullet fl}(j).$$

В рамках каждого из витков с нечетными номерами (coil_1, coil_3, ...) значение абсциссы $X(N_{i,j,k})$ нейрона уменьшается с ростом значения индекса j:

$$X(N_{i,j,k}) = cl \bullet fl - 1 - \mathrm{mod}_{cl \bullet fl}(j).$$

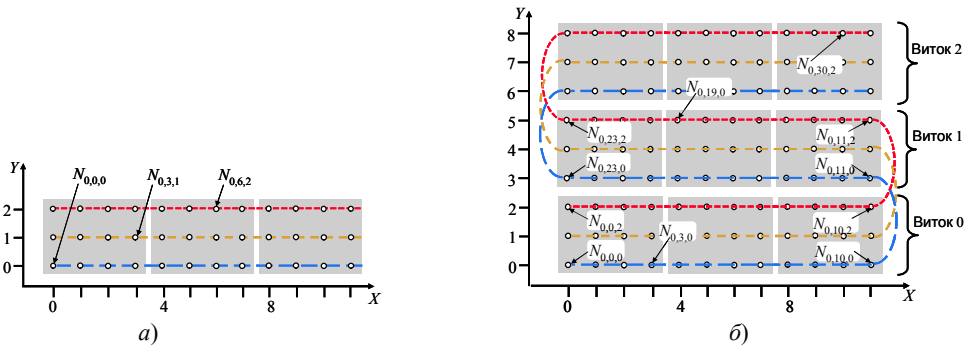


Рис. 3. Варианты структур потоковых РНС:

а) линейная сеть, *б*) многовитковая сеть

Значение ординаты $Y(N_{i,j,k})$ нейрона $N_{i,j,k}$ как на четных, так и на нечетных витках сети определяется выражением

$$Y(N_{i,j,k}) = k + K \bullet \lfloor j /(cl \bullet fl) \rfloor,$$

где $\lfloor x \rfloor$ означает ближайшее к x целое число, не превосходящее x.

Для определения расстояния $D(N_{i,j,k}, N_{i',j',k'})$ может быть использована либо эвклидова метрика $D(N_{i,j,k}, N_{i',j',k'}) = ((j - j')^2 + (k - k')^2)^{-1/2}$, либо, например, метрика Хемминга. Значение фактора β определяется выражением

$$\beta(N_{i,j,k}, N_{i',j',k'}) = h(\alpha \bullet D(N_{i,j,k}, N_{i',j',k'})), \quad 0 < \alpha \le 1,$$

где $h(x)$ — неубывающая функция $0 \le h(x) < 1$, например, гипербола $h = 1 - (1+x)^{-1}$,

Эксперименты с линейной потоковой РНС

Авторами выполнена серия исследований моделей потоковых РНС со значениями параметров $fl = 8$, $K = 5$. В каждом из экспериментов использовались 4 эталона с изображениями букв 'С', 'Е', 'Т', 'Ь' (рис. 4*а*). Эти эталоны размещались во входном модуле последовательно на тактах $t_0, t_4, t_8, t_{12}, \ldots, t_{(4 \bullet lc - 1)}$ где lc — число циклов обучения, $4 \bullet lc$ — общее число экземпляров эталонов, вводимых число в рамках режима обучения.



а) *b*)

Рис. 4. Варианты входных образцов, подаваемых на входной модуль:
а) комплект эталонов, *b*) вариант искажения эталонов

На рис. 4*б* изображены 4 объекта, являющиеся искаженными образцами каждого из эталонов. Искаженные образцы вводятся на тактах $t_{4 \bullet lc}$, $t_{4 \bullet (lc+1)}$, $t_{4 \bullet (lc+2)}$, … и далее до завершения работы сети в режиме восстановления. Если в режиме обучения веса синапсов настроены должным образом, то искаженные образцы в ходе продвижения через сеть восстанавливаются полностью. Представляет интерес наблюдение в режиме обучения за динамикой изменения весов синапсов. Для корректной настройки весов запоминающих синапсов требуется правильный выбор числа lc циклов обучения. При недостаточно большом числе lc циклов обучения веса запоминающих синапсов не успеют возрасти до тех значений, которые обеспечат должное восстановление искаженных образцов эталонов. При чрезмерно большом числе циклов обучения происходит перенасыщение весов запоминающих синапсов и, как следствие, появление в выходном модуле лишних элементов со значениями $o_{j,k}(t) = 1$.

На рис. 5*а* и 5*б* приведены результаты обработки искаженных образцов рис. 4*b* одной и той же потоковой нейронной сетью при недостаточно большом (рис. 5*а*) и чрезмерно большом (рис. 5*б*) числе циклов обучения. При недостаточно большом значении lc искаженный образец 'Е' не восстанавливается полностью. При слишком большом значении lc на выходном модуле возникают лишние элементы $o_{j,k}(t) = 1$ при выводе результата обработки искаженного образца 'С'.



а) *б*)

Рис. 5. Результаты обработки искаженных образцов эталонов из рис. 3*б*:
а) при недостаточном числе циклов обучения, *б*) при чрезмерно большом числе циклов обучения

В ходе наблюдений за динамикой параметров состояния сети в режиме обучения выделены три разновидности запоминающих синаптических связей: шаговые, встречные и протяженные связи.

Три разновидности запоминающих синаптических связей. Пусть для нейрона $N_{i,j,k}$, наряду с запоминающей синаптической связью $(N_{i,j,k}, N_{i',j',k'})$ существует направляющая связь $(N_{i,j,k}, N_{i^*,j^*,k^*})$

такая, что нейроны $N_{i',j',k'}$ и N_{i^*,j^*,k^*} принадлежат одному и тому же полю. Тогда ($N_{i,j,k}$, $N_{i',j',k'}$) относится к разряду *шаговых* синаптических связей. Шаговые синаптические связи играют двоякую роль. С одной стороны, они действуют аналогично направляющим связям в том смысле, что способствуют продвижению обрабатываемых образцов на один шаг. С другой стороны, при достаточном насыщении значений своих весов они способствуют восстановлению недостающих элементов в неполных образцах.

Запоминающая синаптическая связь ($N_{i,j,k}$, $N_{i',j',k'}$) является *встречной* связью, если связь ($N_{i',j',k'}$, $N_{i,j,k}$) является либо шаговой, либо направляющей связью.

Запоминающая синаптическая связь ($N_{i,j,k}$, $N_{i',j',k'}$) является *протяженной* связью, если она не относится ни к разряду шаговых, ни к разряду встречных.

Пачки шаговых синапсов. Для каждой направляющей связи ($N_{i,j,k}$, N_{i^*,j^*,k^*}) число шаговых связей ($N_{i,j,k}$, $N_{i',j',k'}$) равно ($fl \bullet K$) – 1. В секции 2 введено структурное понятие пачки направляющих связей. С этим понятием связано понятие пачки шаговых связей.

Пусть две направляющих связи ($N_{i,j,k}$, $N_{i',j',k'}$) и ($N_{l,x,y}$, $N_{l',x',y'}$) принадлежат пачке по некоторому конкретному эталону. Тогда шаговая связь ($N_{i,j,k}$, $N_{l',x',y'}$) и шаговая связь ($N_{l,x,y}$, $N_{i',j',k'}$) принадлежат *пачке шаговых связей* по этому эталону. Другими словами, пачка шаговых связей объединяет все шаговые связи ($N_{i,j,k}$, $N_{i',j',k'}$), в которых нейрон $N_{i,j,k}$, принадлежит следу конкретного эталона в поле слоя L_i и нейрон $N_{i',j',k'}$ принадлежит следу того же эталона в поле слоя $L_{i'}$.

Число шаговых связей в пачке по конкретному эталону равно $pa \bullet (pa - 1)$, где pa — число ненулевых элементов эталона.

В линейных потоковых РНС пачки шаговых синаптических связей служат главным средством накопления информации об образцах, обрабатываемых в режиме обучения.

Пассивность встречных синаптических связей. При наблюдениях за значениями весов синаптических связей обнаружено, что веса встречных синаптических связей остаются нулевыми на всех этапах работы потоковой РНС по правилам, представленным в разделе 3. После выявления этого факта стало ясно, что эта особенность встречный связей может быть обоснована теоретически. Следовательно, включение встречных синаптических связей в структуры потоковых РНС рассмотренного типа является бесполезным, отказ от них позволяет более рационально расходовать ресурсы памяти при реализации потоковых РНС и их действующих моделей. При этом уменьшается и число действий, требуемых для реализации каждого шага работы сети.

Протяженные синаптические связи. В обоих режимах работы сети обрабатываемые объекты продвигаются по сети от поля к полю в заданном порядке. Если некоторый объект поступает из входного модуля в поле $F_{0,0}$ на такте t_x, то в своем движении по сети он достигает поля $F_{l,j}$ на такте t_{x+l+2j}. Это дает основание рассматривать величину $T(F_{i,j}, F_{i',j'}) = i' + 2j' - i - 2j$ как *тактовое расстояние* между полями $F_{i,j}$ и $F_{i',j'}$. Значение тактового расстояния может быть как положительным, так и отрицательным. Если $T(F_{i,j}, F_{i',j'}) > 0$, то путь от $F_{i,j}$ к $F_{i',j'}$ направлен в будущее, если $T(F_{i,j}, F_{i',j'}) < 0$, то путь от $F_{i,j}$ к $F_{i,j'}$ направлен в прошлое.

Определение расстояния $D(N_{i,j,k}, N_{i',j',k'})$ между нейронами в разделе 3 может рассматриваться как пространственное, геометрическое расстояние. По аналогии с понятием тактового расстояния между полями может рассматриваться другая характеристика относительного размещения нейронов — тактовое расстояние между нейронами $T(N_{l,j,k}, N_{l',j',k'}) = T(F_{l,x}, F_{l',x'})$, где нейрон $N_{i,j,k}$ принадлежит полю $F_{l,x}$, нейрон $N_{l',j',k'}$ принадлежит полю $F_{l',x'}$.

Для синаптических связей, относящихся к разряду направляющих, шаговых или встречных, выполняется равенство $\left| T(N_{l,j,k}, N_{l',j',k'}) \right| = 1$ (абсолютное значение тактового расстояния между передающим и принимающим нейронами равно единице). Для протяженных синаптических связей выполняется неравенство $\left| T(N_{l,j,k}, N_{l',j',k'}) \right| \geq 2$, что и служит основанием называть эти связи протяженными.

Выполненные авторами эксперименты показывают, что в линейных потоковых РНС протяженные связи могут являться причиной зашумления выходных образцов — появления в выходном модуле ложных элементов (элементов со значениями $o_{j,k}(t) = 1$, подобно тому, как в представленном на рис. 4*б* результате обработки искаженного эталона 'C'). Такие эффекты могут вызываться не только перенасыщением весов шаговых синапсов при избыточном числе циклов обучения, но также и нежелательным влиянием протяженных синаптических связяей.

Поэтому при построении линейных потоковых РНС следует избегать использования протяженных синаптических связей.

Специфика работы многовитковых потоковых РНС

Анализ работы многовитковых потоковых РНС показывает, что для них функции ассоциативной памяти наиболее эффективно реализуются межвитковыми протяженными синаптическими связями. связи обеспечивая надежное восстановление искажений входных объектов. Такие синаптические связи обеспечивают надежное восстановление искажений входных объектов даже при отсутствии в многовитковой сети тех самых шаговых синаптических связей, которые в линейных потоковых РНС выполняют роль основной ассоциативной памяти.

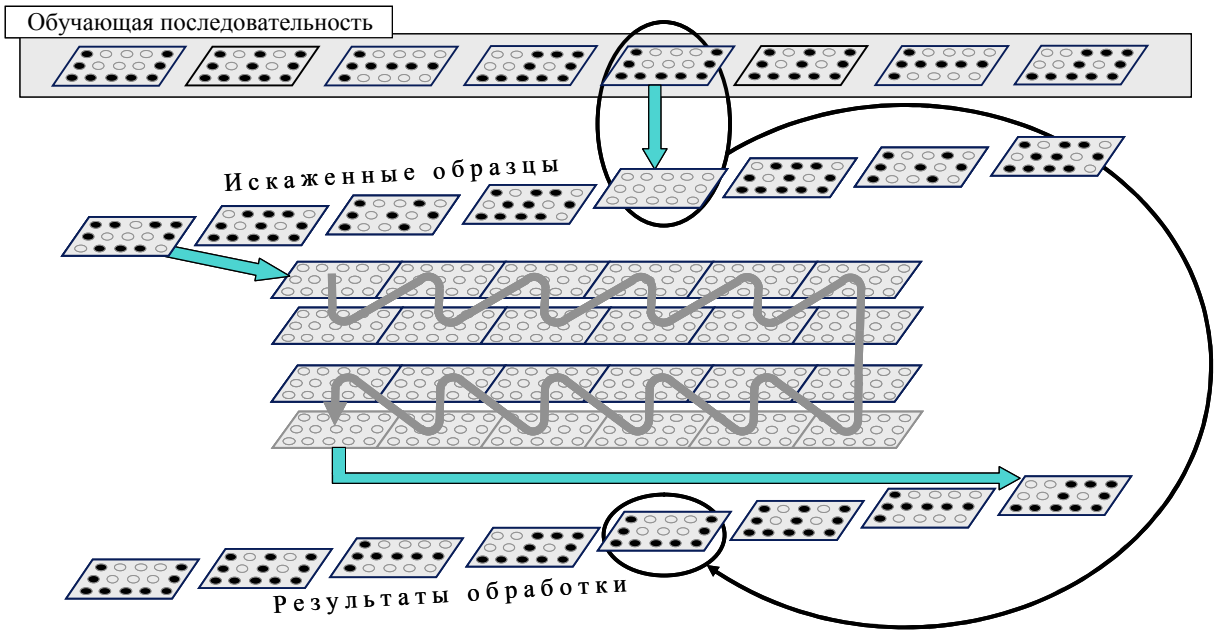


Рис. 6. Восстановление искаженных образцов при использовании межвитковых синапсов

Использование в качестве ассоциативной памяти межвитковых синапсов вместо шаговых синапсов приводит к существенным различиям в характере восстановления искажений эталонных объектов:

Использование шаговых синапсов обеспечивает устранение искажений в каждом отдельном образце входной последовательности. Изменение порядка следования искаженных образцов относительно порядка следования эталонов в режиме обучения не влияет на результаты обработки искаженных образцов.

При использовании межвитковых синапсов порядок следования искаженных образцов должен повторять порядок следования эталонов в режиме обучения. В противном случае в искаженные образцы не будут должным образом восстанавливаться сетью.

На рис. 6 приведен вариант обработки двухвитковой сетью последовательности искаженных образцов, следующих в порядке, повторяющем порядок следования соответствующих эталонов в обучающей последовательности. Образцы входной последовательности содержат искажения двоякого рода:

— элементы со значением 0 при значении 1 соответствующего элемента эталона,

— элементы со значением 1 при значении 0 соответствующего элемента эталона.

Пусть искажения первого рода являются утраченными элементами образца, искажения второго рода - ложными элементами образца. В результате обработки искаженных образцов устраняются искажения и того и другого рода. Первый из последовательности искаженных образцов, соответствующий букве 'С', содержит 2 утраченных элемента ($e_{2,0}$, $e_{2,4}$) и 2 ложных элемента ($e_{0,1}$, $e_{0,3}$). И те и другие элементы успешно восстанавливаются. Активные элементы пято-

го образца утрачены полностью - на входной модуль поступает пустой образец. Тем не менее восстанавливаются все 9 утраченных элементов.

На рис. 7 в последовательности искаженных образцов имеется нарушение порядка их следования относительно порядка следования эталонов в обучающей последовательности (поменялись местами 5ый и 7ой образцы).

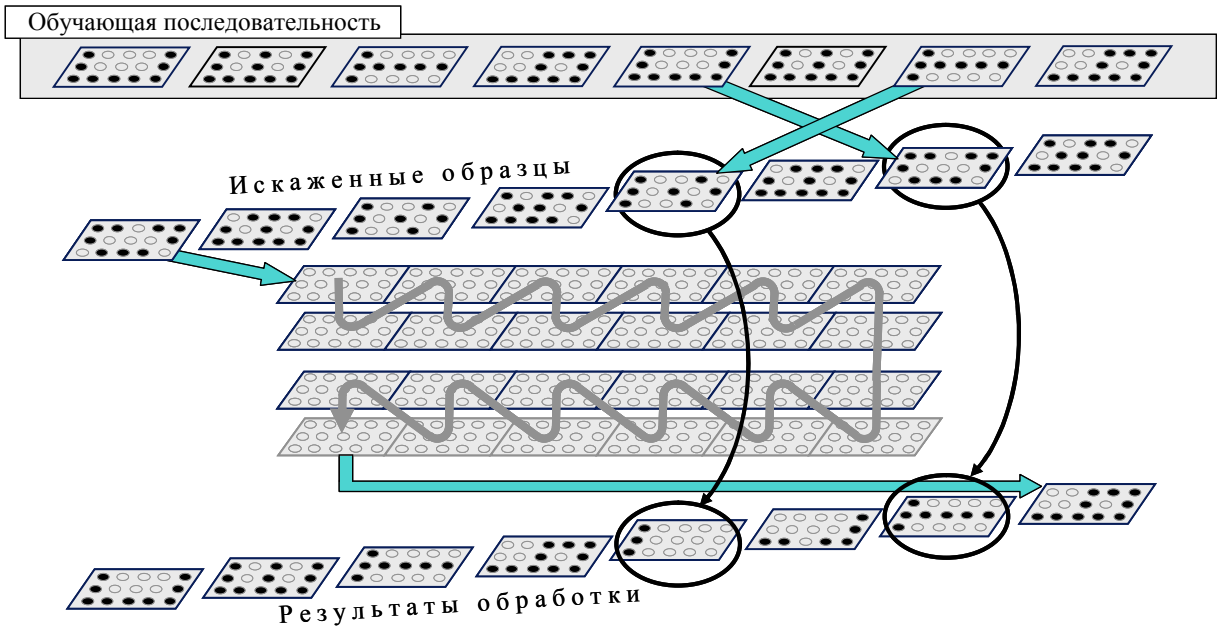


Рис. 7. Нарушение порядка следования искаженных образцов

В результате обработки такой последовательности остаются искаженными не только пятый и седьмой элементы выходной последовательности, но и находящийся между ними шестой элемент.

Заключение

В потоковых РНС синаптические связи используются как элементы ассоциативной памяти. В режиме обучения сетью обрабатывается ряд эталонных объектов с целью настройки ассоциативной памяти на последующее восстановление эталонов по их искаженным образцам. Такое восстановление осуществляется путем реакции ассоциативной памяти на искаженный образец. Выделяются три разновидности запоминающих синаптических связей – шаговые, встречные и протяженные. В линейных потоковых РНС работа ассоциативных механизмов основана в основном на использовании шаговых синаптических связей. В многовитковых потоковых РНС работа ассоциативных механизмов опирается в первую очередь на протяженные межвитковые связи.

В рамках принятых правил функционирования потоковых сетей веса встречных связей остаются нулевыми по завершении режима обучения. На результаты работы сети в режиме восстановления встречные связи влияния не оказывают. Их включение в конструкцию потоковых РНС нецелесообразно.

Использование протяженных связей оказывает полезный эффект только в качестве межвитковых связей в многовитковых сетях. В линейной потоковой РНС и внутри витков многовитковой потоковой РНС использование протяженных связей может приводить к зашумлению выходных образцов.

ЛИТЕРАТУРА

1. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Nat. Acad. Sci*. USA. 1982. Vol. 79, n. 8. P. 2554–2558.
2. Kosko, B.:. Bidirectional associative memories. *IEEE Trans. Syst. Man Cybern*. 1982. Vol. 18, n. 1. P. 49–60.
3. **Osipov, V.Yu.** Space-Time Structures of Recurrent Networks with Controlled Synapses. *Proceedings of 13th International Symposium on Neural Networks – ISNN 2016*. St-Petersburg, Russia, 2016. P.177-184.
4. **Osipov, V.Yu.**, **Nikiforov, V.V.** Formal Aspects of Streaming Recurrent Neural Networks. . *Proceedings of 15th International Symposium on Neural Networks – ISNN 2016*. Minsk, Belarus, 2018. P.29-36.

V.Yu. Osipov, V.V. Nikiforov (St.Petersburg Institute for Informatics and Automation of Russian Academy of Sciences, St.Petersburg)

Streaming recurrent neural networks with guided elements

The specific features of streaming recurrent neural networks with linear and multi-coil structures and guided elements are regarded. The rules are defined for associative processing of signals by such neural networks. Each signal (image) has the form of a sequence of single pulses. The contribution is analyzed of various synapses kinds in such associative processing of signals. The results of modeling with simultaneous restoration of lost elements and suppression of superfluous elements are presented. The ways for decreasing of structural complexity of considered neural networks and increasing the ability of their associative memory are founded due to results of realized modeling.

ИССЛЕДОВАНИЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ УПРАВЛЕНИЯ АВТОНОМНОГО НАЗЕМНОГО РОБОТА С НЕЙРОННОЙ СЕТЬЮ ГЛУБОКОГО ОБУЧЕНИЯ

В. Х. Пшихопов[1], **М. Ю. Медведев**[2], **В. Б. Пархоменко**[3], **М. А. Васильева**[4]
(*НИИ робототехники и процессов управления ЮФУ, Таганрог, ул. Шевченко, 2, medvmihal@sfedu.ru*)

Аннотация

Рассматривается колесный робот, состояние которого определяется линейными координатами и углом ориентации, а управлениями являются скорости вращения колес. На основе анализа кинематики предложена классификация траекторий в пространстве скоростей вращения колес. Нейронная сеть используется для распознавания типа траектории. Проведенное исследование показало, что для достижения заданной вероятности правильной классификации требуется большая выборка изображений. В этой связи рассмотрена система управления с идентификатором, реализованным нейронной сетью, которая осуществляет идентификацию параметров робота. Вначале система управления осуществляет выполнение нескольких маневров, позволяющих определить параметры робота с требуемой вероятностью. Количество маневров определяется заданной вероятностью правильной классификации. Показано, что в системе с идентификатором снижены требования к объему выборки для обучения нейронной сети. Также приводятся результаты численного моделирования системы управления с нейронными сетями глубокого обучения.

Ключевые слова: мобильный робот, нейросетевое управление.

Введение

Первой сетью глубокого обучения, успешно примененной для решения задачи распознавания символов, является сверточная нейронная сеть [1]. Сети глубокого обучения широкого используются в задачах распознавания текстов [2] и объектов [3], навигации [4], понимания сцен [5], переноса выученных закономерностей в другие области [6]. Перспективными направлениями нейросетевых систем являются исследования, связанные с технологиями глубокого обучения и самообучения с подкреплением, с использованием баз знаний (онтологий) и программ логического вывода, оперативного дообучения.

В системах автоматического управления нейронные сети используются в качестве адаптивных регуляторов, для планирования движений и идентификации объекта. При этом особое внимание занимают вопросы обеспечения устойчивости и обучения нейронных сетей.

В работе [7] рассматривается адаптивная нейросетевая следящая система управления вертолетом при наличии параметрических и функциональных неопределенностей. Нейросетевое управление синтезировано с использованием метода Ляпунова, что позволяет гарантировать слежение за опорными сигналами с малыми ошибками.

Вопросы обеспечения устойчивости и обучения нейронный сетей в системах автоматического управления также рассмотрены в работе [8].

В [9] представлен алгоритм обучения для динамических рекуррентных нейронных сетей Эльмана, основанный на оптимизации частиц Роя. Разработан новый метод управления, в котором динамический идентификатор выполняет идентификацию скорости ультразвукового двигателя.

В [10] приводятся нейросетевые алгоритмы планирования движения в двумерной неопределенной среде. В результате сравнительного анализа показана высокая эффективность нейронных сетей при решении задач планирования пути.

В [11] предложена нейронная сеть для идентификации обратной динамики дискретного объекта, что позволяет придать системе управления прогнозирующие свойства.

Применение нейронных сетей в качестве регуляторов движения требует высокой степени надежности, которая достигается при объемах обучающих выборок в десятки миллионов изображений. В данном докладе исследуется система управления наземным колесным роботом, со-

[1] Доктор технических наук, директор.
[2] Доктор технических наук, ведущий научный сотрудник.
[3] Техник.
[4] Техник.

зданная с использованием сети глубокого обучения. При этом выбирается такая структура нейросети, которая позволяет пользоваться уменьшенной обучающей выборкой.

Постановка задачи

Рассматриваются уравнения кинематики мобильного робота с двумя ведущими колесами с управляемой скоростью вращения [12, 13]:

$$\begin{aligned} \dot{x} &= r_k\left(\omega_R + \omega_L\right)\cos\varphi, \\ \dot{y} &= r_k\left(\omega_R + \omega_L\right)\sin\varphi, \\ \dot{\varphi} &= r_k\left(\omega_R - \omega_L\right)/a, \end{aligned} \tag{1}$$

где x, y – линейные координаты робота в неподвижной системе; φ – угол рысканья робота; ω_R, ω_L – скорости вращения левого и правого колеса; r_k, a – радиус колес и ширина робота.

С целью упрощения изложения, взята упрощенная модель робота. Однако представленные далее результаты могут быть перенесены на более сложные модели, т.к. предлагаемые алгоритмы нейросетевой идентификации не базируются на модели робота.

Требуется разработать систему управления, обеспечивающую перевод робота из произвольной начальной точки с координатами x, y в целевую точку, заданную координатами x_c, y_c. Управляющими входами робота являются скорости вращения колес ω_R, ω_L. Ставится задача разработать такие алгоритмы изменения скоростей вращения ω_R, ω_L колес робота, которые базируются на процедуре обучения, использующей изображения траекторий движения. В результате обучения должны быть сформированы функции $\omega_R\left(x, y, \varphi\right)$, $\omega_L\left(x, y, \varphi\right)$. При этом модель (1) не используется при формировании указанных фукций. Исходные изображения траекторий формируются эталонной системой или в результате экспериментальных измерений движений робота.

Система управления с нейросетевым классификатором траекторий движения

Рассмотрим систему прямого нейросетевого управления [14], в которой нейронная сеть является регулятором движения.

Для обучения нейросетевого регулятора синтезируется эталонная система управления, обеспечивающая перевод робота, описываемого уравнениями кинематики (1), из произвольной начальной точки с координатами x, y в целевую точку, заданную координатами x_c, y_c.

Введем направляющий вектор вида

$$p_n = \begin{bmatrix} x_c - x & y_c - y \end{bmatrix}^T.$$

Потребуем, чтобы скорость движения к целевой точке была максимальной при большом расстоянии до нее и уменьшалась при приближении к ней. В частности, такая зависимость может быть задана соотношением вида

$$V = V_{\max} sat\left(\gamma_r \left\| p_n \right\|\right), \tag{2}$$

где V – текущая линейная скорость робота; $V_{\max}$ – максимальная скорость; γ_r – постоянный коэффициент, обратно пропорциональный расстоянию до целевой точки, на котором мобильный робот начинает торможение; $\left\| p_n \right\|$ – Евклидова нормы; $sat\left(\left\| p_n \right\|\right)$ – функция насыщения, ограничивающая норму значениями от -1 до +1.

Желаемый угол рыскания робота определяется таким образом, чтобы он двигался на целевую точку, т.е.

$$\varphi_c = \arctan\left(\frac{y_c - y}{x_c - x}\right). \tag{3}$$

При этом требования к текущему углу рысканья, с учетом последнего выражения, определяются соотношением

$$\dot{e}_\varphi + T_\varphi e_\varphi = 0,$$

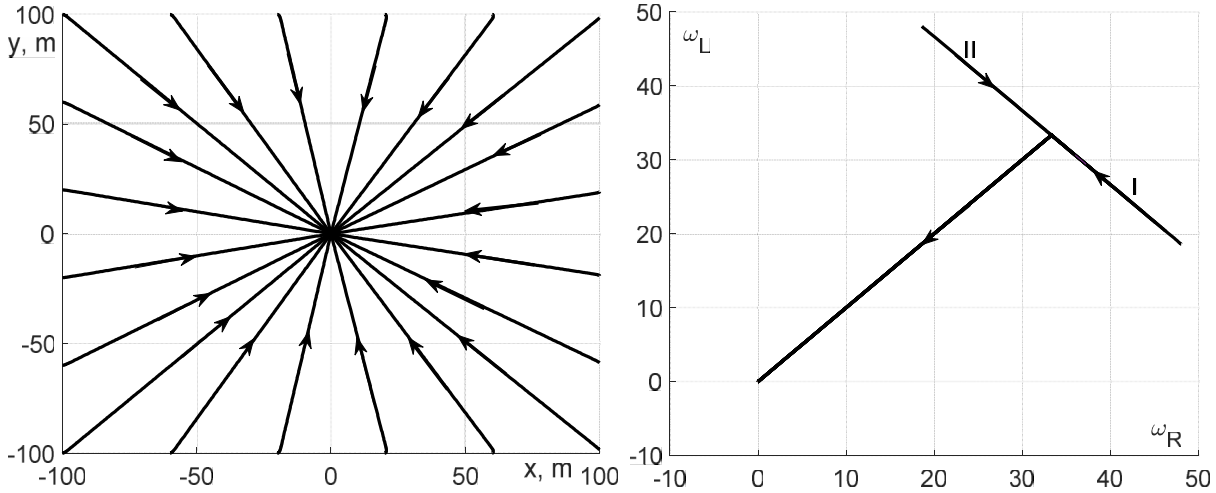
где $e_\varphi = \varphi_c - \varphi$ – ошибка по углу рысканья; T_φ – постоянный коэффициент, определяющий скорость разворота.

С учетом последнего выражения, модели кинематики (1) и требований к скорости и углу (2), (3), получаем выражение для вычисления управляющих воздействий (скоростей вращения колес робота)

$$\begin{bmatrix} \omega_R \\ \omega_L \end{bmatrix} = \frac{1}{r_k} \begin{bmatrix} 1 & 1 \\ 1/a & 1/a \end{bmatrix}^{-1} \begin{bmatrix} V \\ -T_\varphi(\varphi_c - \varphi) \end{bmatrix}. \quad (4)$$

Результаты моделирования замкнутой системы управления (1) – (4) при различных начальных условиях представлены на рис. 1. Моделирование проводилось при следующих параметрах: $T_\varphi = 3$; $V_{\max} = 2$; $x_c = 0$, $y_c = 0$; $\gamma_r = 0.5$; $r_k = 0.03$, $a = 0.1$.

На рис. 1а представлены траектории движения робота в целевую точку при различных начальных условиях, а на рис. 1б – траектории движения в плоскости скоростей вращения колес ω_R, ω_L. Как видно из представленных результатов моделирования, все траектории движения можно разделить на два класса, обозначенные как I и II на рис. 1б. Эти два класса являются пересекающимися, т.е. содержат одинаковый элемент. Моделирование при других координатах целевой точки и при различных углах рысканья также показывают, что на плоскости ω_R, ω_L в замкнутой системе существует только два типа траекторий. Отличие заключается только в том, что при различных начальных условиях замкнутая система стартует из различных точек либо с траектории I, либо с траектории II.



а) Траектории движения б) Плоскость скоростей колес

Рис. 1. Результаты моделирования системы управления робота

В процессе обучения нейронной сети предъявляются изображения траекторий движения, которые она учится классифицировать на принадлежность к классу I или II. В результате обучения для каждой траектории, соединяющей две точки на плоскости *Oxy*, формируется функция $\omega_L^I(\omega_R)$ или $\omega_L^{II}(\omega_R)$. Причем эти функции состоят из двух прямых. Стартовая точка на прямой I или II выбирается исходя из начальных условий.

Этапы работы обученного нейронного регулятора заключаются в следующем. 1) По начальной $x(0)$, $y(0)$ и конечной x_c, y_c точкам выбирается изображение эталонной траектории робота. Эта траектория классифицируется обученной нейронной сетью, т.е. выбирается функция $\omega_L^I(\omega_R)$ или $\omega_L^{II}(\omega_R)$. 2) Исходя из начальных условий $x(0)$, $y(0)$ вычисляется начальная точка $\omega_L(0)$, $\omega_R(0)$, принадлежащая функции $\omega_L^I(\omega_R)$ или $\omega_L^{II}(\omega_R)$. 3) С заданным интервалом времени осуществляется коррекция текущей начальной точки x, y.

С использованием среды MATLAB, инструментов Neural Network Toolbox, спроектирована сверточная нейронная сеть, структура которой представлена на рис. 2.

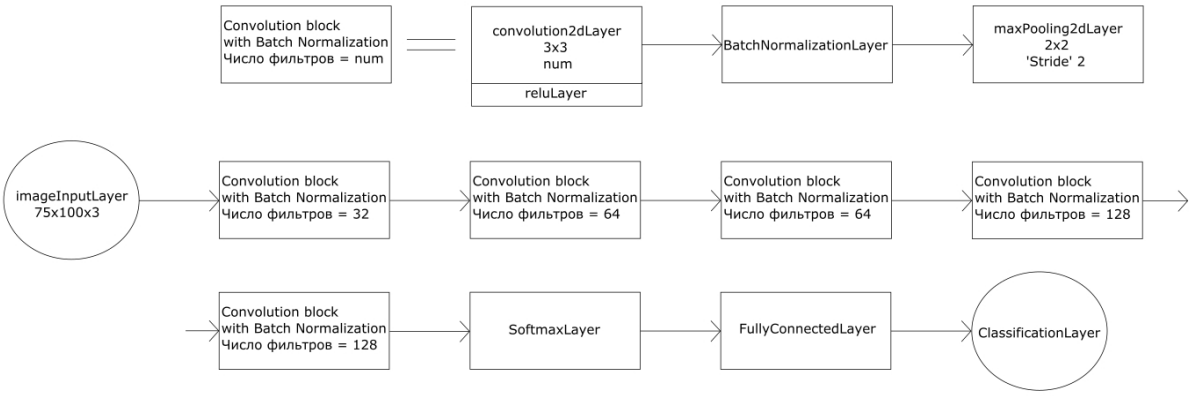


Рис. 2. Структура спроектированной сверточной нейронной сети

Данные на входной слой imageInputLayer подаются в виде JPG файла размерностью 75x100 пикселей, и разбиваются на 3 канала. Слой convolution2dLayer является скользящим фильтром с ядром размерностью 3×3 пикселя, который осуществляет свертку входа. После свертки применяется слой batchNormalizationLayer, в котором используется выпрямленная линейная функция активации, существенно повышающая сходимость градиентных методов обучения [15]. После каждого сверточного слоя, функции активации и слоя нормализации, изображение обрабатывается фильтром maxPooling2dLayer, размерностью 2x2, отбирающим максимальные значения из матрицы значений пикселей. Слой SoftmaxLayer применяется для того, чтобы сеть можно было использовать для классификации по числу категорий, более двух. Например, такая ситуация возникает при наличии реверса скорости вращения колес. Выход слоя Softmax подается на полносвязный слой fullyConnectedLayer, число нейронов которого соответствует числу классов. Этот слой умножает входы на весовые коэффициенты и добавляет вектор смещения, что позволяет оптимизировать нейронную сеть. В выходном слое classificationLayer окончательно определяется искомый класс.

В ходе обучения нейронной сети осуществлена оптимизация размера пакетов, на которые разбиваются данные при обучении и количество прохождений всех полученных пакетов через нейронную сеть. Наибольшая полученная точность распознавания проверочного набора данных составила 86%. Результат обучения представлен на рис. 3.

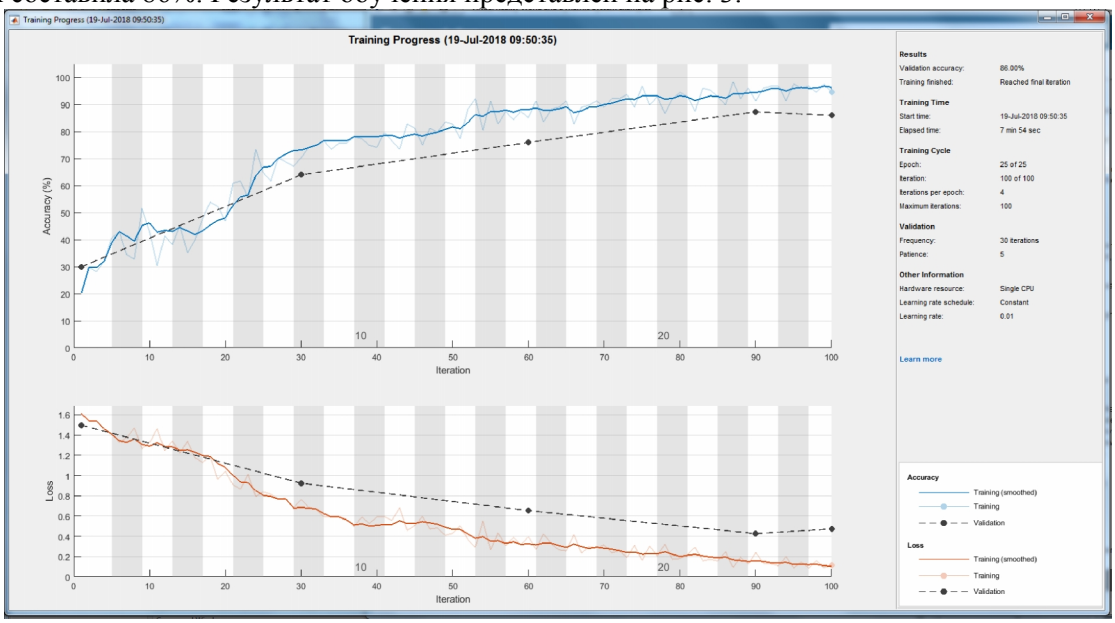


Рис. 3. Результат обучения нейронной сети

Полученный результат не позволяет реализовать надежную систему управления движением. В этой связи требуется существенное увеличение обучающей выборки и связанное с ним повышение качества изображения. Альтернативой данном способу повышения качества управления является использование системы управления с нейронным идентификатором.

Система управления с нейросетевым идентификатором

Задачей нейронного идентификатора является классификация колесного робота по принадлежности к одному из классов, представленных в табл. 1.

Таблица 1

Платформы для классификации

Класс	Радиус колес, см	Ширина платформы, см
1	1	2
2	1	3
3	1	4
4	2	4

Исходными данными для нейронной сети являются изображения траекторий движения робота, полученные при различных скоростях вращения левого и правого колеса и различных начальных условиях.

Разработка системы управления мобильного робота с нейросетевым идентификатором описывается следующими этапами.

1) Создание генератора траекторий движения робота при различных скоростях вращения колес и начальных условиях.

2) Генерация необходимого числа изображений и последующее их сжатие до необходимого разрешения.

3) Обучение нейронной сети.

Генератор изображений представляет программный код а Matlab, позволяющий автоматически получить необходимое число траекторий при различных начальных условиях и скоростях вращения колес. Автоматизация этого процесса существенно сокращает время получения обучающей выборки. С помощью разработанного генератора получено 1280 изображений, по 320 для каждого класса, представленного в табл. 1. Скорости вращения ведущих колес робота изменялись от 1 до 14 рад/с. Примеры полученных изображений для обучения нейронной сети представлены на рис. 4.



Рис. 4. Изображения для обучения нейросетевого идентификатора

Полученные изображения далее сжимались до разрешения 200×150. Сжатые изображения использовались для обучения, в процессе которого достигнута точность правильной классификация 75 %.

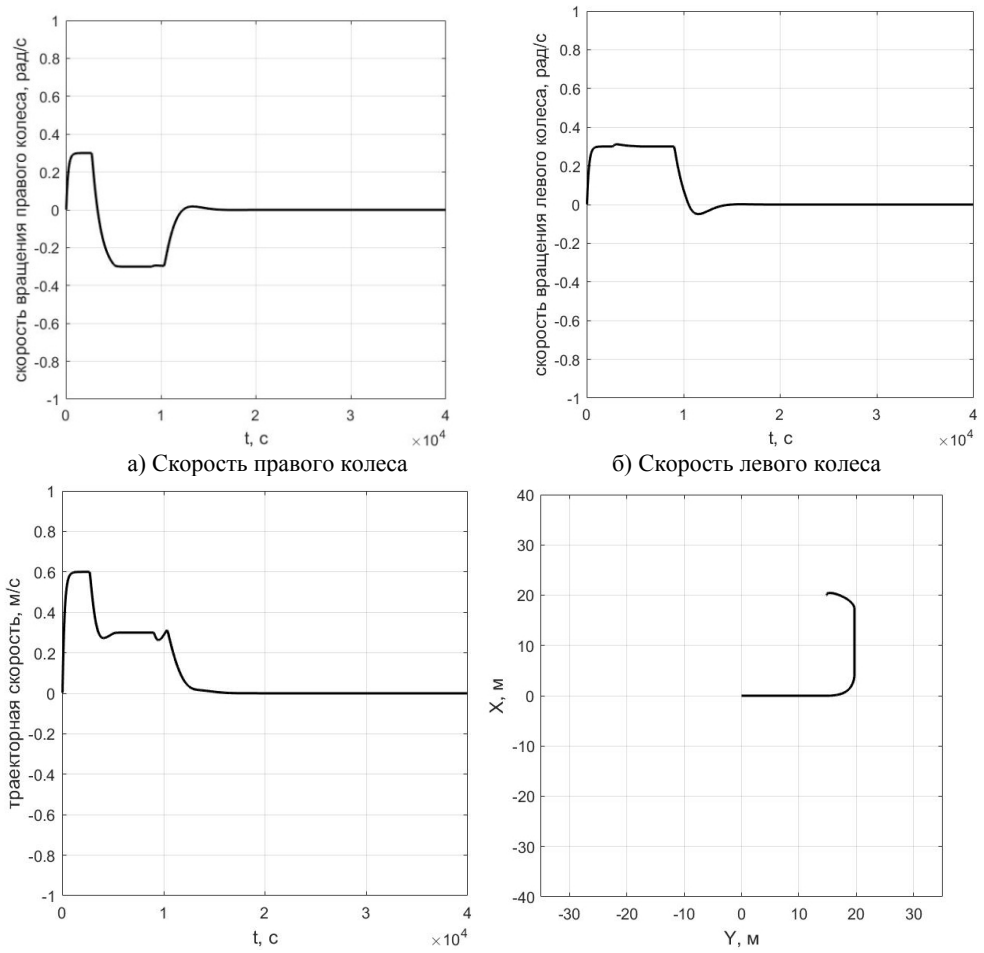
Перед началом функционирования системы управления проводится активный эксперимент, состоящий из n опытов. Число опытов n рассчитывается по формуле Бернулли [16]

$$P_{n,k} = \sum_{k=1}^{n} C_n^k p^k q^{n-k}, \quad C_n^k = \frac{n!}{k!(n-k)!},$$

где p – вероятность правильной классификации в каждом испытании; $q = 1 - p$; k – количество случаев правильной классификации.

Из последнего выражения следует, что при n=20, вероятность правильной идентификации робота составляет 0,986.

На рис. 5 представлены результаты моделирования системы управления с нейросетевым идентификатором для робота класса 4.



а) Скорость правого колеса б) Скорость левого колеса

в) Траекторная скорость робота б) Траектория движения робота

Рис. 5. Результаты моделирования системы управления с нейросетевым идентификатором

Результаты моделирования показали, что ошибки идентификации не приводят к нарушению устойчивости траекторий движения, однако перерегулирование и длительность переходных процессов отклоняются от эталонных значений на 10 – 20 %.

Заключение

В докладе представлены подходы к разработке интеллектуальных систем управления наземных роботов, отличающиеся тем, что управляющие нейросети обучаются по изображениям траекторий движения. Это позволяет в регуляторах движения применять без всяких изменений нейронные сети и методы их обучения, используемые в системах распознавания изображения.

Полученные в докладе результаты демонстрируют перспективы использования нейронных сетей в качестве идентификаторов моделей подвижных объектов. В частности, для наземных роботов в процессе функционирования, как правило, требуется уточнять только коэффициенты продольного и поперечного скольжения. Остальные параметры определяются с достаточной точностью на этапе создания модели объекта. Аналогично, для автономных необитаемых подводных аппаратов наиболее неопределенными параметрами являются гидродинамические коэффициенты [17], а для летательных аппаратов – аэродинамические коэффициенты [18]. Проблема повышения точности идентификации может решаться каскадным применением нейросетей [19]. При таком подходе первая нейросеть классифицирует идентифицируемый параметр на принадлежность к одному из двух интервалов. Далее, каждый интервал делится пополам, а вторая нейросеть осуществляет классификацию внутри каждого интервала. Такой подход упрощает обучение нейронной сети и позволяет с достаточной точностью идентифицировать требуемые параметры.

Работа выполнена при поддержке гранта РНФ № 18-19-00621

ЛИТЕРАТУРА

1. LeCun Y., Boser B., Denker J.S., Henderson D., Howard R. E., Hubbard W., and Jackel L.D. Backpropagation Applied to Handwritten Zip Code Recognition, *Neural Computation*, 1989. N 1(4). P. 541-551.
2. Kim Y. Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, (2014). P. 1746–1751.
3. Angelova A., Krizhevsky A., Vanhoucke V. Pedestrian detection with a Large-Field-Of-View deep network. *IEEE International Conference on Robotics and Automation*. 2015.
4. Engel J., Schöps T., Cremers D. LSD-SLAM: Large-Scale Direct Monocular SLAM. *Lecture Notes in Computer Science*, 2014. Vol 8690.
5. Zhu J., Park T., Phillip Isola, and Alexei A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", in IEEE International Conference on Computer Vision (ICCV), 2017.
6. Gholami B., Rudovic O., Pavlovic V. Probabilistic Unsupervised Domain Adaptation for Knowledge Transfer Across Visual Categories. *The IEEE International Conference on Computer Vision (ICCV)*, 2017, Pp. 3581-3590.
7. Ge S., Ren B., and Tee K. Adaptive neural network control of helicopters with unknown dynamics. *45th IEEE Conference on Decision and Control*. 2006. P. 3022 – 3027.
8. Терехов В.А., Ефимов Д.В., Тюкин И.Ю. Нейросетевые системы управления. М.: Высшая школа, 2002. 183 с.
9. Hong-Wei G., Wen-Li D., Feng Q., Lu W. A Dissimilation Particle Swarm Optimization-Based Elman Network and Applications for Identifying and Controlling Ultrasonic Motors. *Advances in Soft Computing*. 2007. Vol 41. P. 1345-1360. Springer.
10. Pshikhopov, V.Kh. (Ed.), Beloglazov D., Finaev V., Guzik V., Kosenko E., Krukhmalev V., Medvedev M., Pereverzev V., Pyavchenko A., Saprykin R., Shapovalov I., Soloviev V. Path Planning for Vehicles Operating in Uncertain 2D Environments, 2017. Elsevier, Butterworth-Heinemann, 312p.
11. Ноткин Б.С. Эффективная нейросетевая идентификация инверсной динамики объекта управления для синтеза прогнозирующих систем управления. *Труды IV Междунар. конф. Идентификация и задачи управления SICPRO'05*. 2005.
12. Буданов В.М., Девянин Е.А. О движении колёсных роботов. *Прикладная математика и механика*. 2003, Т. 67. Вып. 2. С. 244 – 255.
13. Мартыненко Ю.Г. Управление движением мобильных колесных роботов. *Фундаментальная и прикладная математика*. 2005. No8. С. 29 – 80.
14. Омату С., Халид М., Юсоф Р. Нейрокомпьютеры и их применение. М.: Радиотехника. 2000.
15. Krizhevsky A., Sutskever I., and Hinton G.E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*. 2012.
16. Вентцель Е.С. Теория вероятностей: Учебник для вузов. 5-е издание. М.: Высшая школа. 1998. 576 с.
17. Пшихопов В.Х., Суконкин С.Я., Нагучев Д.Ш., Стракович В.В., Медведев М.Ю., Гуренко Б.В., Костюков В.А., Волощенко Ю.П. Автономный подводный аппарат «Скат» для решения задач поиска и обнаружения заиленных объектов. *Известия ЮФУ. Технические науки*. 2010. – № 3(104). С. 153 – 162.
18. Pshikhopov, V., Medvedev, M., Gaiduk, A., Neydorf, R. , Belyaev, V., Fedorenko, R., Krukhmalev, V. Mathematical model of robot on base of airship. *2013 Proceedings of the IEEE Conference on Decision and Control*. P. 959-964.
19. Angelova A., Krizhevsky A., Vanhoucke V. et al. Real-Time Pedestrian Detection With Deep Network Cascades. *British Machine Vision Conference*. 2015.

V.Kh. Pshikhopov, M.Yu. Medvedev, V.B. Parkhomenko, M.A. Vasil'eva (Southern Federal Unirsity, Taganrog)

Development of control system of wheeled mobile robot on base of deep learning neural network

The control system of a wheel robot is considered. The model of the robot is described by kinematics equations. The controls are the speeds of the wheels. Based on the analysis of the kinematics, the classification of trajectories in the state space of wheel rotation speeds is proposed. The neural network is used to recognize the trajectory type. The study showed that to achieve the required probability of correct classification of the available sample is not enough. In this regard the control system with identification is considered. Identification algorithm is based on neural network. The neural network performs the identification of the parameters of the robot. Initially, the control system performs an experiment. The experiment consists of several maneuvers that allow determine a robot parameters with the required probability. The number of maneuvers is determined by the given probability of correct classification. It is shown that in the system with the identifier the requirements to the sample size for neural network training are reduced. The results of numerical simulation of the control system with deep learning neural networks are also carried out.

ОБЛАЧНАЯ ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА SMART FARMING ДЛЯ УПРАВЛЕНИЯ ТОЧНЫМ ЗЕМЛЕДЕЛИЕМ

П. О. Скобелев[1], **В. Б. Ларюхин**[2],
(*Самарский государственный технический университет,
443100, г. Самара, ул. Молодогвардейская, 244, Россия,
petr.pkobelev@gmail.com; vl@smartsolutions-123.ru*)

Е. В. Симонова[3]
(*Самарский национальный исследовательский университет им. академика С.П. Королёва,
443086, Самара, Московское шоссе, д.34, Россия,
simonova@smartsolutions-123.ru*)

Д. С. Будаев[4], **Г. Ю. Вощук**[5],
(*ООО «НПК «Разумные решения»,
443013, Самара, ул. Московское шоссе, 17, офис 12-01, Россия,
budaev@smartsolutions-123.ru; voschuk@smartsolutions-123.ru*)

Аннотация

Рассматривается система SMART FARMING, предназначенная для интеллектуализации управления предприятиями растениеводства, занимающимися разработкой и внедрением технологий точного земледелия.

Описана архитектура системы на основе цифровой Интернет-платформы I-Can-Farm, включающей в себя ряд «умных сервисов», которые позволяют автоматизировать процессы принятия решений в хозяйствах, прежде всего, в части непрерывного мониторинга полей и оперативного управления техникой, а также расчета экономики в реальном времени.

Особенностью системы является применение баз знаний и мультиагентных технологий для выработки согласованных решений по распределению, планированию, оптимизации и контролю ресурсов предприятия в реальном времени.

Ключевые слова: точное земледелие, интеллектуальная система управления, адаптивное планирование, агент, мультиагентные технологии

Введение

Сельское хозяйство составляет значимую долю (4,6%) в объеме ВВП России и играет ключевую роль в экономике страны. По оценке аналитической компании J`son & Partners Consulting, Россия занимает 3-е место в мире по площади пашни, выращивает 8,4% от мирового урожая пшеницы и занимает 1-е место по экспорту пшеницы (более 40 млн. т. в 2017 г.). Вместе с тем, для сохранения позиций РФ на мировом рынке необходимо повышать качество и эффективность производства, переработки и хранения продукции.

Современная тенденция развития сельского хозяйства обусловлена активным внедрением в производство развитых стран точных ресурсосберегающих методов земледелия, включающих производственные технологии, высокотехнологичные производственные агрегаты и информационные технологии [1]. Одним из базовых элементов ресурсосберегающих технологий в сельском хозяйстве является «точное земледелие» (precision agriculture) – комплексная высокотехнологичная система сельскохозяйственного менеджмента, целью которой является получение максимальной прибыли при условии оптимизации сельскохозяйственного производства, экономии хозяйственных и природных ресурсов [2].

Предлагаемый доклад посвящен рассмотрению облачной интеллектуальной системы Smart Farming – универсального программного решения на базе мультисервисной платформы I-Can-Farm для согласованного управления сельскохозяйственным производством растениеводческо-

[1] Доктор технических наук, заведующий кафедрой.
[2] Аспирант.
[3] Кандидат технических наук, доцент
[4] Руководитель департамента
[5] Программист.

го профиля с применением баз знаний и мультиагентных технологий с целью выработки согласованных решений по распределению, планированию, оптимизации и контролю ресурсов предприятия в реальном времени.

Система Smart Farming позволяет перевести «в цифру» основные ресурсы предприятия растениеводства (поля, технику и другие), создать его «цифровой двойник» и вести основные производственные операции с помощью современных Интернет-технологий. Система Smart Farming рассматривается как базовая цифровая Интернет-платформа для дальнейшего наращивания «умными сервисами», которые позволят автоматизировать процессы принятия решений в хозяйствах, прежде всего, в части непрерывного мониторинга полей и оперативного управления техникой, а также расчета экономики и бизнес-аналитики предприятий в реальном времени.

Особенности управления растениеводческой отраслью сельского хозяйства

Для эффективного управления растениеводством следует учитывать следующее важные факторы:

1. Процесс выращивания – это адаптивный процесс, требующий быстрой, точной и гибкой реакции на возникающие непредвиденные события, при этом стратегический и последующие тактические планы работы хозяйства на основе имеющихся мощностей (финансовых ресурсов, парка техники, квалификации рабочих и т.д.), многократно будут изменяться в ходе сезона.

2. Технология – это набор типовых, отработанных и выработанных операций, которые сложились из опыта, экспертных знаний, научных исследований и т.п. Технология включает также рекомендуемые способы реакции на возникающие непредвиденные события – сорняки, вредителей, болезни и прочее. Знания об эффективности операций постоянно видоизменяются, пополняются и накапливаются. Технология заключается в том, чтобы найти ответ на вопросы, что делать в тех или иных ситуациях, чем подкармливать растения или противостоять вредителям и т.п. Это означает, что жесткое программирование даже самой передовой технологии в предлагаемой платформе нецелесообразно, т.к. она может измениться и потеряет свою актуальность для хозяйств.

3. Главными факторами в развитии культур и применении технологий являются: влага, минералы (состав почвы), сорт культуры, сорта культур-предшественников, погодные условия. Для культур можно задать начальный профиль (и далее формировать уточняемый профиль), который подскажет, какие компоненты, в каком количестве она требует, как будет развиваться при различных показателях.

4. По каждому полю важно иметь информацию о химическом балансе, который говорит о том, сколько пришло элементов (осадков, минералов – из удобрений) и сколько потрачено культурой – в будущем может быть построена модель почвы.

5. Важная цель системы – собрать достаточную аналитику, чтобы далее разбираться и делать выводы, почему не получен ожидаемый результат, или какие мероприятия привели к результату выше ожидаемого: сроки их проведения, эффект, контекст (погода, культура, сорт в привязке к стадии развития растений).

6. Не только сама технология возделывания культур постоянно развивается, но и специалистами проводятся эксперименты с удобрениями, сортами растений, подходами к обработке и т.п., которые позволяют получать более высокие урожаи, требующие более интенсивного труда, четкого выполнения сроков и применения дорогой техники, что не всегда получается в хозяйствах на практике, нужно понимать критический уровень допустимых отклонений.

7. Необходимо сравнивать технологии, применяемые для разных полей, чтобы понимать их отличительные особенности, анализировать реакцию на возникающие события и конечный результат. Необходимо накопление такой информации о процессах и фактах деятельности, без этого невозможно выполнять аналитику или строить базы знаний, что очень важно для всех хозяйств.

8. Необходимо автоматизировать процесс управления работами механизаторов в сельском хозяйстве, чтобы система могла строить планы использования техники и подсказывать, какие вводить работы, визуализировать и контролировать результаты работы каждой бригады механизаторов.

9. Мониторинг и контроль техники для определения затрат горюче-смазочных материалов (ГСМ) должен заключаться не только в учете передвижений и местоположения техники, но и в фиксации времени выполнения отдельных этапов операций – сколько прошло времени от постановки задачи до выезда бригады, время в пути до поля, время работы в поле, скорость передвижения по полю, время отъезда с поля.

10. На основании собираемой информации об операциях можно видеть отклонения от средних значений с учетом видов работ и площадей полей, и оценивать положительные результаты для мотивации сотрудников.

11. Система мотивации должна быть направлена на повышение производительности и качества результатов труда, например, путем связывания расчета премий с полученной урожайностью культур на каждом поле, на котором работали агрономы и механизаторы, а также с устранением простоев и затрат вследствие не выполненных вовремя ремонтов и профилактического обслуживания техники.

Планирование в растениеводстве имеет ряд специфических особенностей, обусловленных необходимостью комплексно учитывать значительно б*о*льшую по сравнению с другими отраслями совокупность факторов:

- биологических, к которым относятся биологические циклы производства, защита растительного мира, карантинные правила, агроклиматические условия, видовые и сортовые требования и т.п.;
- технологических – в сельском хозяйстве технологии чрезвычайно дифференцированы и диверсифицированы. Они зависят от природно-экономических условий, технической оснащенности производства, финансовых возможностей производителя. Варианты технологий требуют основательной экспертизы и анализа их экономической эффективности. Надежным инструментом такой оценки являются технологические карты и разработанные на их основе нормативы расходов;
- технических, что обусловлено необходимостью применять в сельском хозяйстве широкий ассортимент технических средств. Большинство из них используется в течение ограниченного промежутка времени. Технические средства часто не унифицированы, эксплуатационные и расходные материалы, запасные части не являются взаимозаменяемыми. Поэтому стоимость техники является относительно высокой по сравнению с исходным продуктом;
- экологических – природная среда является неотъемлемой составляющей аграрного производства, поэтому соблюдение правил экологической безопасности является обязательным. Производственные процессы не могут ставить под угрозу экологическую безопасность региона, местности, конкретного ландшафта. Экологичность хозяйственной деятельности улучшает социальный имидж сельхозпредприятия, облегчает сотрудничество с общественностью, местными органами власти и является экономически выгодной;
- экономических – экономическая эффективность аграрного производства в мире ниже, чем в других отраслях производства, оно дотируется. Получение результата является пролонгированным во времени и имеет большой уровень риска.

Таким образом, выращивание растений методически следует рассматривать как процесс управления сложными адаптивными системами [3], требующий принятия решений всеми участниками в консенсусе, учитывающем особенности почвы, погоды, сортов растений, наличие ресурсов и т.д.

Цифровая платформа и эко-система сервисов на основе мультиаегнтных технологий

В результате анализа производственных процессов на сельскохозяйственных предприятиях предложено создавать платформу как открытую эко-систему «умных сервисов» («систему систем»), где каждый сервис, в свою очередь, представляет собой автономную систему, способную принимать решения и взаимодействовать с другими такими же системами (Рис. 1).

В экосистеме параллельно и асинхронно работающих «умных сервисов» каждый сервис представлен программным агентом, способным к взаимодействию с другими агентами сервисов через общую шину предприятия. Создаваемые сервисы проектируются таким образом, чтобы

сохранять возможность взаимодействовать с другими внешними системами и встраиваться в другие платформы.



Рис. 1. Конструкция платформы как экосистемы умных сервисов

Основные функции системы позволяют оперативно реагировать на непредвиденные внешние события, адаптивно планировать работу подразделений предприятия и контролировать использование ресурсов в реальном времени:

1. Онлайн-база знаний по растениеводству, включающая важные сведения о культурах растений, почвах, удобрениях, машинах и технике, семенах и др.;

2. «Умное ведение проекта года» – создание сквозного плана работы предприятия на весь год с анализом имеющихся ресурсов в подразделениях и оценкой рисков;

3. «Умный мониторинг» – агент каждого поля не просто запланирует и предоставит изображение поля для анализа хода роста и развития растений или контроля результатов каждой важной операции, но и проведет автоматическое выявление и классификацию проблемных ситуаций на вверенном поле с уведомлением агрономов на сотовый телефон о возможном типе обнаруженной проблемной ситуации;

4. «Умные сотрудники» – каждый сотрудник предприятия получает агента, который строит оперативные планы (сменно-суточные задания) на день и адаптивно меняет их по возникающим непредвиденным событиям;

5. «Умная техника» – каждое важное событие будет вызывать адаптивное перепланирование маршрутов и графиков работы техники с одновременным контролем «план против факта» для каждой единицы техники в реальном времени (бригадирам и механизаторам теперь не надо вручную перестраивать планы и строить сменно-суточные задания на каждый день, причем сами задачи в этих заданиях могут меняться в течение дня по ходу работы в полях);

6. «Умные финансы» – агент финансиста каждого поля оценит возможные затраты и доходы по каждому полю и по каждой культуре (против котлового метода), выявит возможные проблемы и подскажет решения, а также будет далее перестраивать финансовые прогнозы прибыли в реальном времени.

Для создания платформы предлагается использовать мультиагентные технологии и базы знаний на основе онтологий [4, 5].

Разработанная система имеет веб-серверную архитектуру и работает через все основные браузеры типа Internet Explorer, Google Chrome и другие, интегрируется через API или передачу данных с другими системами (1С и др.).

Метод виртуального круглого стола для поддержки принятия согласованных решений агентами

Ниже приведены принципы проектирования агентов:

1. Каждый агент сервиса рассматривается как машина состояний – это означает, что агент, как и мультиагентная система в целом, работает в цикле автономного функционирования, реагирует на события, принимает решения, согласовывает эти решения с другими агентами и

добивается их реализации. По внешним командам можно инициализировать любого агента сервиса, запросить и проверить его состояние, попросить завершить работу и т.д.

2. Каждый программный агент построен на основе универсальной конструкции, получающей информацию о событиях во внешнем мире, вырабатывающей решения и согласующей эти решения с другими агентами. Конструкция базового агента включает базовый системный блок управления, не зависящий от предметной области, и специализированный для предметной области блок, что позволяет быстро настраивать агента на сервис в любой предметной области, встраивая в него содержательные модели, методы и алгоритмы.

3. Базовый агент представляет собой программную конструкцию, построенную как машина состояний (по типу конечного автомата с памятью), переходящую из состояния в состояние под действием событий, в процессе анализа данных предприятия или поступления сообщений, причем при переходе в новое состояние контекст агента сохраняется в хранилище данных.

4. Каждый из агентов «материализуется» по запросу нового пользователя, либо от агентов поступившей задачи или другого сервиса, а также ресурса предприятия. Для реализации требований задачи агент остается активным, пока функции не будут исполнены, он обладает еще и внутренней активностью, постоянно стремится достичь заданной ему цели и выполняет все требуемые текущие задачи до получения результата.

В результате исследований для управления предприятием растениеводства в развитие концепции ПВ-сетей [6] предложен следующий состав базовых классов агентов (Таблица 1).

Таблица 1

Классы агентов

Тип агента	Функции и цели
Агент поля	Определяет выбор наиболее подходящей культуры, чтобы добиться эффективности поля, с учетом растений-предшественников, типов почвы, рельефа и других факторов.
Агент культуры	Определяет выбор наиболее подходящего поля для сорта растения и далее – порядок операций возделывания и контроля по техкартам процессов, генерирует задачи для агентов техники, бригад, мониторинга.
Агент средства защиты растений (СЗР)	Определяет тип СЗР для внесения на поле. Планирует технологические операции по внесению СЗР. Минимизирует объем внесения пестицидов.
Агент удобрений	Планирует задачи от агронома по типам, порядку и срокам внесения удобрений. Минимизирует объем внесения удобрений в соответствии с техпроцессом.
Агент мониторинга	Определяет выбор спутников или дронов для отработки запросов и возможные сроки доставки изображений, предоставляет данные по съемке полей, определяет индексы поля, выявляет неоднородности на полях.
Агент вредителей (насекомых) или болезней	На основе выявленных неоднородностей на полях и смежных параметров (погодные условия, календарь развития болезней, близость лесного участка и т.п.) определяет наиболее вероятных вредителей или возможные болезни растений, а также тип возможных СЗР для лечения или борьбы с насекомыми.
Агент агронома	Планирует севооборот и определяет культуры для посевов на полях, контролирует состояние полей, определяет технологические операции для каждого поля и растения.
Агент бригады	Формирует, согласовывает и контролирует выполнение плана по обработке посевов (план сменно-суточных заданий).

Агент техники	Формирует план обработки полей, задает требования по техобслуживанию, подбирает по типам специализации механизаторов и технику, минимизирует время простоев.
Агент механизатора	Планирует рабочее время, календарь занятости по согласованию с агентом агронома.
Агент штабной (модератор)	Является модератором «круглого стола», контролирует заданные агрономом показатели и поддерживает процесс согласования планов агентами.

Модель виртуального «круглого стола» предполагает возможность создания и применения различных методов и алгоритмов выработки и принятия согласованных коллективных решений. При этом вместо реальных специалистов-экспертов от их лица и по поручению действуют программные агенты агро-сервисов (Рис. 2):

- агент спутников идентифицирует, что одно из полей СХ предприятия – проблемное;
- агент дронов планирует уточненную съемку поля;
- агенты заболеваний растений и агенты вредных насекомых определяют в качестве причины высокую вероятность наличия вредителей;
- агент агронома планирует выезд агронома в поле для принятия решений на месте, в результате обнаруживается, что причина другая – недокорм растений;
- вводится событие «недокорм растений», агент удобрений планирует сервис внесения удобрений с учетом имеющихся в наличии или быстро доставляемых удобрений;
- агенты техники и агенты механизаторов планируют новые сменно-суточные задания и отправляют их на планшеты механизаторов.

По завершению процесса внесения удобрений поле ставится на мониторинг и контроль, чтобы отследить изменения и добиться получения результата.

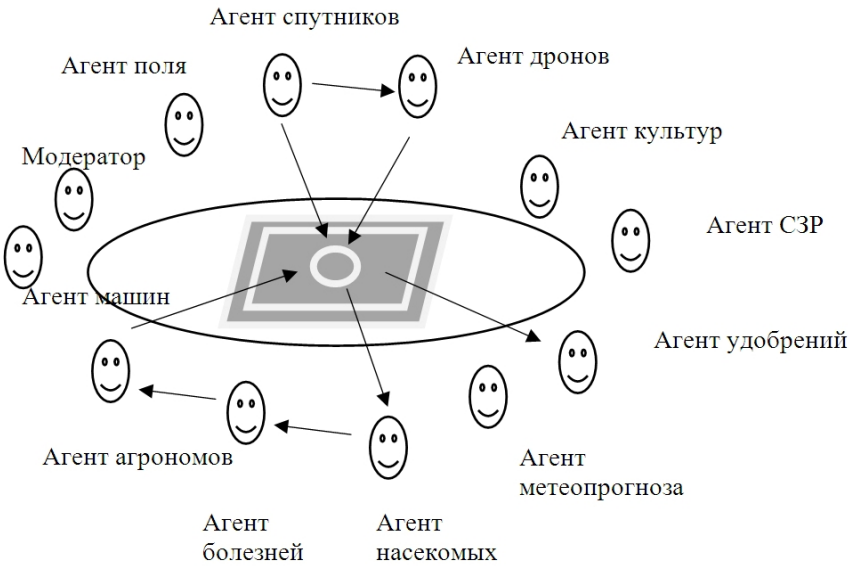


Рис. 2. Пример виртуального «круглого стола» для выработки согласованных решений

Общая архитектура виртуального «круглого стола» включает следующие компоненты:

1. Базу знаний, содержащую основные роли и регламенты переговоров агентов сервисов и других участников виртуального «круглого «стола» для принятия согласованных решений, описание задач и свойств культур;

2. Агента-модератора (штабного агента), управляющего ходом ведения переговоров, включая вызов нужных агентов-представителей сервисов, выбор регламента переговоров, инициализацию нового раунда переговоров, передачу очереди между участниками, отработку очередного хода, фиксацию результата, выдачу результата;

3. Агентов-представителей сервисов, которые «знают» общее состояние своих сервисов, но при необходимости могут отправлять своим сервисам запросы на проработку тех или иных событий, одновременно порождая копию текущего состояния сервиса;

4. Сцену «круглого стола», на которой размещается формализованная модель данных о проблемной ситуации, требующей разрешения, план действий, выработанный путем переговоров, и результат переговоров агентов для решения проблемы.

Наиболее универсальным методом согласования решений является метод переговоров «по кругу» с возвратами.

Разработка базы знаний для управления точным земледелием

Онтологический подход к хранению знаний позволяет с помощью онтологий описать любую разнородную, сколь угодно сложную предметную область. С точки зрения разработки интеллектуальных систем, используя этот подход, можно формализовать специфические предметные знания в виде, допускающем компьютерную обработку, и отделить знания от программного кода системы. Возможность для пользователей онтологии добавлять и изменять концепты, атрибуты и объекты по мере уточнения целей и задач делает этот метод пригодным в практическом использовании.

Таким образом, онтологии на базовом уровне должны, прежде всего, обеспечивать словарь понятий (терминов, концептов) для представления и обмена знаниями о предметной области, а также множество связей (отношений), установленных между понятиями в этом словаре [7].

До определения понятий и отношений предметной области растениеводства разрабатывается набор родительских базовых классов и отношений. Понятие «Сущность» является родительским классом для всех других описываемых классов, на основе которого разработаны следующие базовые классы: «Заказ», «Задача», «Ресурс», «Технология», «Результат». Для каждого из базовых классов необходим набор атрибутов, которые описывают присущие данному классу уникальные характеристики. Базовые понятия через отношение наследования могут быть конкретизированы производными понятиями.

База знаний системы SmartFarming содержит коллекцию специализированных онтологий, описывающих различные аспекты производства в аграрной отрасли сельского хозяйства:

1. Онтология растениеводства, содержащая понятия (классы) для описания сельскохозяйственных культур: «Культура», «Группа», «Сорт», «Гибрид», «Характеристика», «Посевной материал», «Фаза роста», «Потребность», «Посев», «Состояние», «Продукт». В дальнейшем при создании онтологии предметной области возможно расширение разработанных понятий (классов). Например, понятие «Характеристика» для описания сельскохозяйственной культуры может быть расширено дополнительными атрибутами или использовано как родительское понятие (класс) для описания новых понятий: «Структура», «Индекс урожая» (отношение между зерном и соломой), «Опасность переуплотнения почвы и подпочвы», «Содержание аминокислот» и другие.

2. Онтология болезней посевов, содержащая понятия для описания болезней посевов, насекомых-вредителей посевов: «Фактор», «Вирусное заболевание», «Грибковое заболевание», «Вредитель», «Симптом», «Компонент влияния», «Возбудитель-источник».

3. Онтология почв, содержащая понятия для описания типов и параметров почв: «Почва», «Вид почвы», «Характеристика», «Состояние». Класс «Характеристика» фактически представляет собой тип для описания критериев оценки качества почв, позволяющий описать показатели водного режима, гумусового состояния, засорения, гранулометрический состав, скелетность, pH-фактор, засоленность, агроклиматическую группу и другие параметры почв. Класс «Состояние» разработан для описания фактического состояния почв той или иной агроклиматической группы по всем параметрам, применяемых для описания критериев оценки качества почв.

4. Онтология технологических процессов, содержащая понятия для описания технологических процессов (карт) обработки посевов: «Технологическая карта», «Севооборот», «Тип операции», «Погодные условия», «Облачность», «Солнечная радиация», «Температура», «Влажность». «Давление», «Ветер». Технологическая карта в агропроизводстве – это документ, содержащий необходимые сведения о проводимых операциях, используемых ресурсах, инструкциях для персонала, выполняющего технологические операции по возделыванию сельскохозяйственных культур. Операции в технологической карте обычно представлены в порядке их пла-

нируемого выполнения во времени, в том числе, указываются календарные сроки, тип сельскохозяйственной операции, объем (площадь) работ, затраты ресурсов.

Кроме типов понятий (классов) для описания технологических процессов были разработаны понятия для описания погодных условий, поскольку погодные условия являются ограничивающим факторов для проведения сельскохозяйственных операций.

5. Онтология удобрений и средств защиты растений, содержащая понятия для описания средств защиты растений: «Средство защиты растения», «Удобрение», «Фактор действия», «Тип действия на фактор», «Группа действующих веществ», «Действующее вещество».

6. Онтология сельскохозяйственных машин, содержащая понятия для описания машин и оборудования: «Механизм», «Сельхозмашина», «Тип техники», «Характеристика», «Техническое состояние», «Операция обслуживания». Понятие «Характеристика» предполагается использовать для создания других дочерних понятий (классов), описывающих параметры техники.

7. Онтология организации, содержащая понятия для описания кадрового состава, должностей, компетенций: «Сотрудник», «Должность», «Профессия», «Время», «Подразделение», «Организация».

8. Онтология товарно-материальных ресурсов, содержащая понятия для описания товарно-материальных ресурсов: «Вещество», «Инфраструктура» и др.

Кроме понятий (классов) разработан набор отношений, применимых для связи одних понятий (классов) с другими: «Является», «Входит в», «Состоит из», «Предшествует», «Следует за», «Требует», «Потребляет», «Производит», «Относится к».

Систематизация сведений о хозяйственных объектах и процессах предприятия в виде семантически связанной базы знаний делает их пригодными для компьютерной обработки с учетом всей полноты особенностей и взаимосвязей между ними (Рис. 3).

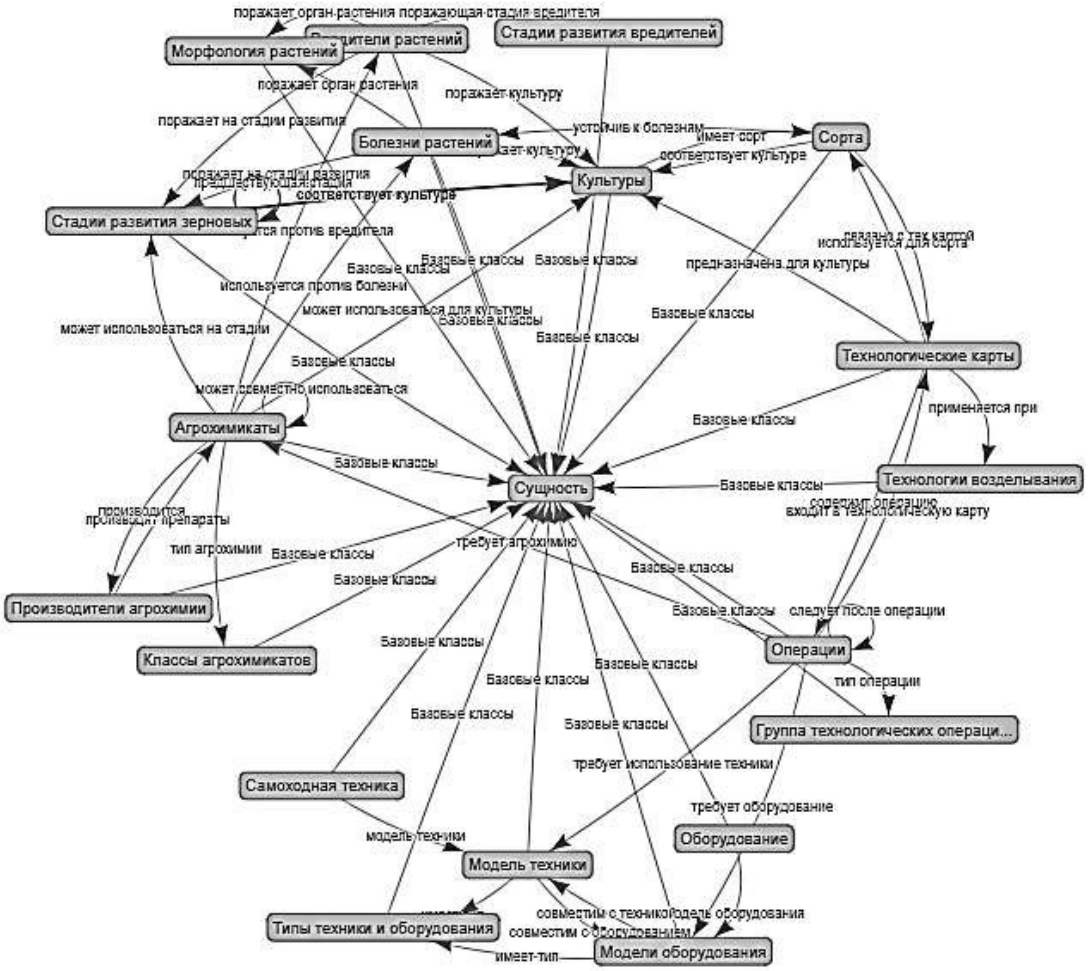


Рис. 3. Онтология растениеводства в виде семантической сети

Таким образом, на основе онтологии могут обеспечиваться, например, проверка готовности материальной базы предприятия к внедрению новых технологий обработки почв и посевов, проверка соответствия сортов и культур особенностям климатических зон, автоматический подбор предпочтительных удобрений и агрохимикатов для использования в конкретных сложившихся условиях, с учетом текущего состояния посевов, проблем на полях, фаз развития растений. Интеллектуальная программная система Smart Farming предоставляет для автоматической обработки структурированные и связанные данные, с которыми могут одновременно работать как специалисты по управлению сельскохозяйственными предприятиями, так и работники аграрной отрасли, путем внесения в базу новых знаний, полученных в результате практической работы на полях (сорта, культуры, удобрения, технологии обработки почв, технику, указывая значения их атрибутов и взаимосвязи между объектами и процессами).

Заключение

Преимущества использования интеллектуальной системы Smart Farming для решения задачи управления предприятиями растениеводческого профиля:

- Сокращение сложности и трудоемкости в управлении ресурсами;
- Налаживание учета и контроля в использовании ресурсов;
- Полная 24/7 прозрачность предприятия для собственника и руководства;
- Все данные предприятия хранятся в одном, доступном для всех сотрудников месте в защищенном облаке;
- Удобство, наглядность и простота в использовании для специалистов предприятия;
- Веб-интерфейс для дистанционного доступа из любых географически удаленных точек.
- Перспективы развития:
- новые модели управления предприятиями, ориентированные на результаты;
- компьютерные модели развития растений для создания киберфизических систем (контроль 100 фаз развития пшеницы и т.п.);
- аналитика больших данных и машинного обучения на базе нейронных сетей.

Работа проводилась при поддержке гранта Министерства образования и науки Российской Федерации – договор № 14.574.21.0183, уникальный идентификационный номер RFMEFI57417X0183.

ЛИТЕРАТУРА

1. Агропромышленный комплекс России в 2016 году. Росинформагротех, 2017. 704 с.
2. Рунов Б.А., Пильникова Н.В. Основы технологий точного земледелия: зарубежный и отечественный опыт. Спб: Рос. акад. с.-х. наук, Гос. науч. учреждение «Агрофизический науч.-исслед. ин-т Россельхозакадемии», 2012, 120 с.
3. Pinedo M. Scheduling: Theory, Algorithms, and Systems. Springer, 2015. 670 p.
4. Skobelev P. Multi-Agent Systems for Real Time Adaptive Resource Management. In Industrial Agents: Emerging Applications of Software Agents in Industry. Paulo Leitão, Stamatis Karnouskos (Ed.). Elsevier, 2015. P. 207–230.
5. Skobelev P. Towards Autonomous AI Systems for Resource Management: Applications in Industry and Lessons Learned / Y. Demazeau et al. (Eds.) // Proceedings of the 16th International Conference on Practical Applications of Agents and Multiagent Systems (PAAMS 2018), Toledo, Spain, 20-22 June, 2018. Advances in Practical Applications of Agents, Multi-Agent Systems, and Complexity. LNAI 10978. Springer, Switzerland. P. 12-25.
6. Виттих В. А., Скобелев П. О. Мультиагентные модели взаимодействия для построения сетей потребностей и возможностей в открытых системах. *Автоматика и Телемеханика*. 2003. №1. С. 177–185.
7. Скобелев П.О. Онтология деятельности для ситуационного управления предприятиями в реальном времени. *Онтология проектирования*. 2012. № 1(3). С. 6–38.

P.O. Skobelev (Samara State Technical University, Samara), **E.V. Simonova** (Samara National Research University, Samara), **D.S. Budaev** (SEC «Smart Solutions», Ltd., Samara), **G.Yu. Voschuk** (SEC «Smart Solutions », Ltd., Samara), V.B Larukhin (Samara State Technical University, Samara)

Smart farming: cloud system for precise agriculture.

SMART FARMING system is designed for intellectualization of management in plant-growing enterprises, engaged in development and introduction of precise agriculture technologies.

The paper describes system architecture based on the digital Internet-platform I-Can-Farm, which includes a number of "smart services" making it possible to automate decision-making processes in farms, primarily in continuous field monitoring and operational machinery management, as well as cost analysis in real time.

One of the distinguishing features of the system is using the knowledge base and multi-agent technology for coordinated decisions on allocation, planning, optimization and control of resources of plant-growing enterprises in real time.

ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ НЕЙРОННЫХ СЕТЕЙ ДЛЯ УПРАВЛЕНИЯ СИСТЕМОЙ САМОДИАГНОСТИКИ СУДОВОЙ ЭНЕРГЕТИЧЕСКОЙ УСТАНОВКИ

В. Е. Колпаков[1]
(*ВМПИ ВУНЦ ВМФ «Военно-морская академия»,*
196602, С.-Петербург, Пушкин, Кадетский бульвар, 1, val-kolpakov@mail.ru)

Аннотация

В статье приведен анализ систем самодиагностики силовых установок современных поршневых двигателей внутреннего сгорания, позволяющих выявлять неисправность на ранней стадии ее проявления. Рассмотрены наиболее перспективные методы распознавания образов; основное внимание уделено возможности использования нейронных сетей в качестве инструмента самодиагностики, определяющего вероятную причину и место возникновения неисправности. Определена методология синтеза архитектуры нейронной сети «с учителем» для решения задачи создания и управления системой самодиагностики. Указаны основные трудности, возникающие в рамках поставленной задачи.

Ключевые слова: система самодиагностики; система управления двигателем; нейронные сети; диагностические признаки.

Введение

Современный научно-технический прогресс приводит к постоянному усложнению технических систем и, как следствие, к увеличению трудовых и материальных затрат, связанных с эксплуатацией оборудования. В этих условиях широкое внедрение систем самодиагностирования становится одним из важнейших факторов повышения эффективности использования энергетических установок.

Системы самодиагностики двигателей судовых энергетических установок, несмотря на общие подходы, имеют некоторые отличия от систем самодиагностики высокооборотных автотракторных двигателей, что связано, прежде всего, с большими габаритами и высокой стоимостью. При этом для оценки технического состояния производится анализ тенденций изменения диагностических параметров во времени и устанавливается необходимый срок переборки узла. Выход из строя узла предупреждается сигнализацией о достижении предельного значения параметра. Однако, системы самодиагностики в настоящее время не достигли такого состояния, при котором можно было бы полностью положиться на их информацию. Так, системой диагностики не охвачены механические повреждения двигателя или побочные эффекты, порождаемые утечкой вакуума или неисправностью одного или нескольких элементов системы газораспределения и пр., которые могут вызвать проблемы с составом рабочей смеси или холостым ходом.

Предлагаемый доклад посвящен анализу существующих классических систем самодиагностики и оценке перспектив использования нейронных сетей (искусственного интеллекта).

Анализ классических систем управления и самодиагностики силовых установок современных поршневых двигателей внутреннего сгорания и перспективы использования нейронных сетей

Разработка систем управления двигателем и самодиагностики нового поколения является сложной задачей, целью которой является повышение надежности и долговечности силовой установки, снижение трудовых и материальных затрат на ее обслуживание и ремонт. При этом необходимо использование современных диагностических средств и совершенствование математического аппарата определения состояния диагностируемого объекта.

Первая система управления двигателем, известная как Bosch Motronic. была разработана и установлена на автомобиль BMW 732i в 1978году.

Идея системы управления состоит в том, что установленный блок электронного управления (БЭУ) постоянно следит за режимом работы двигателя и подстраивает его параметры так, чтобы в любой момент двигатель работал наилучшим образом. В дальнейшем в БЭУ были введены функции самодиагностики, которые дали возможность водителю или автомеханику следить за

[1] Доктор технических наук, профессор.

состоянием двигателя и определять возникшие неисправности, которые трудно идентифицировать иным образом. Это было достигнуто за счет оснащения БЭУ микропроцессором с памятью и базой данных. Теперь информация о неисправностях может быть сохранена в памяти компьютера и извлечена из нее по мере необходимости. На некоторых автомобилях система самодиагностики извещает водителя о неисправности, на других она может сообщать код неисправности в виде серии вспышек лампочки. Впервые система самодиагностики Bendix Digital была установлена в 1981 году на автомобиле Cadillac. Интенсивное развитие систем управления двигателем в течение 80-х годов привело к тому, что БЭУ, которыми теперь оснащаются современные автомобили, лишь отдаленно напоминают ранние прототипы [1].

В БЭУ введена адаптивная функция, которая может подстраивать базу данных и программу управления в соответствии с текущим состоянием двигателя. Основная идея самодиагностики заключается в проверке соответствия уровня сигналов БЭУ их эталонным аналогам, заложенным в память. Если уровень сигнала выходит за допустимые пределы, БЭУ трактует это как неисправность и заносит в память специальное сообщение. Эти сообщения могут быть вызваны из памяти в виде "кода неисправности". После извлечения эти коды дают важную информацию для диагностики. В современных системах самодиагностики мониторинговые функции существенно расширены и включают в себя анализ и управление процессом сгорания на основе использования одного или нескольких датчиков кислорода, вторичной воздушной системы, системы улавливания паров газов и системы рециркуляции отработавших газов; определение степени сгорания углеводородов и сигнализацию в случае отсутствия возможности обеспечения полного сгорания. Компоненты полного и неполного процесса сгорания представлены на рисунке 1.

Системы самодиагностики двигателей судовых энергетических установок, несмотря на общие подходы, имеют некоторые отличия от высокооборотных и относительно малогабаритных автотракторных двигателей. Так система самодиагностики (CEDC) судовых дизелей фирмы «Зульцер» (Швейцария) на базе мини-ЭВМ анализирует текущие параметры дизеля и его топливной системы. При изменении рабочих диагностических параметров производится анализ тенденций их изменения во времени и устанавливается необходимый срок переборки узла. Выход из строя узла предупреждается сигнализацией о достижении предельного значения параметра. Система самодиагностики судовых дизелей (СС – 10) фирмы «Бурмейстер и Вайн», установленная на дизеле 5ДКРН 62/140-3 теплохода «Александр Прокофьев» обеспечивает контроль работы турбокомпрессора, топливной аппаратуры, цилиндропоршневой группы, воздухоохладителя, охладителей воды, масляных насосов, вспомогательных систем главного двигателя. Однако опыт эксплуатации показал ее малую эффективность и низкую надежность. Система самодиагностики (DETS) фирмы «Норконтрол» (Норвегия) представляет собой измерительный комплекс, дающий информацию о системе впрыска топлива и процесса сгорания в дизеле в виде распечатанных значений параметров и графиков. Система PED фирмы «Пилстик» (Франция) впервые была установлена на дизеле 18PC2-5V теплохода «Ренор». Системой диагностируются среднеоборотный дизель фирмы «СЕМТ-Пилстик» серий РС-2, РС-3, РС-4. Диагностическая система контролирует состояние рамных подшипников коленчатого вала и верхнего поршневого кольца и агрегатов топливной аппаратуры [2].

Мониторинг дизельных двигателей также должен обеспечивать систему диагностики необходимой информацией, но перечень контролируемых элементов в них может быть иным (например, запальные свечи).

ISO, SAE и неправительственные экологические организации выступают за дальнейшее ужесточение правил. Организация "Чистый воздух" (США) внесла в правительство стандарт CARB как основу сохранения окружающей среды и здоровья. Появление каталитических преобразователей, новых систем впрыска топлива, применение топлива без содержания свинца, увеличение относительного числа автомобилей, оснащенных дизельными двигателями, позволили за последние 30 лет достичь значительных позитивных сдвигов в решении проблемы токсичности выхлопных газов.



Рисунок 1. Компоненты полного и неполного сгорания топлива

С другой стороны, побочные эффекты, порождаемые, например, утечкой вакуума или неисправностью одного или нескольких элементов системы газораспределения могут вызвать проблемы с составом смеси или холостым ходом, которые приведут к появлению соответствующих кодов. Таким образом, при появлении таких кодов придется проверить многие системы двигателя, чтобы обнаружить истинную причину неисправности, поскольку код указывает только на неисправную цепь. Например, код, указывающий на неисправность цепи датчика температуры охлаждающей жидкости, может означать неисправность самого датчика, либо связанных с ним проводов, либо электрических разъемов.

Таким образом, основной принцип современных систем самодиагностики основан на считывании показателей датчиков, сравнением их показаний с эталонными значениями и в некоторых системах отслеживание скорости изменения аномальных диагностических параметров с последующим прогнозированием ресурса узла или агрегата. Однако, как показывает практика наличие одновременно двух или нескольких неисправностей может ввести такую систему в заблуждение. Так, например: поздний угол впрыска топлива увеличит температуру выпускных газов вследствие растягивания процесса сгорания на линию расширения, а завышенное давление впрыска форсунки снизит цикловую подачу, что приведет к снижению температуры выпускных газов. В итоге температура выпускных газов может оказаться в пределах эталонных значений, и т.п. Кроме того тенденция увеличения количества датчиков требует проведения более сложного анализа их показателей.

Разработка систем управления двигателем и самодиагностики нового поколения является непростой задачей, включающей в себя как применение современных диагностических средств, так и совершенствование математического аппарата определения состояния диагностируемого объекта. Современные системы технической диагностики должны быть непосредственно связаны с распознаванием образа состояния объекта. Первые работы в области теории распознавания и классификации по прецедентам появились в 30-х годах прошлого столетия и были связаны с байесовской теорией принятия решений (работы Неймана, Пирсона), применением разделяющих функций к задаче классификации (Фишер 1936г.), решением вопросов проверки гипотез (Вальд 1939г.). В 50-х годах появились первые нейросетевые модели распознавания (перцептрон Розенблата 1965 г.), связанные с успехами в моделировании процессов головного мозга. К концу 60-х годов уже были разработаны и детально исследованы различные подходы для решения задач распознавания в рамках статистических, перцептронных моделей, и моделей с разделяющими функциями. В настоящее время нейронные сети получили широкое развитие и имеют исключительные возможности для получения значений, правил и тенденций из дан-

ных, затруднительных для понимания, имеющих значительные погрешности и шум. Нейронные сети, управляемые посредством сложных математических функций, могут использоваться для получения образов и распознавания тенденций и решать задачи такой сложности, которые не под силу для математических моделей, использующих аналитические или параметрические методики. Одним из достоинств нейронных сетей является способность точного прогнозирования значений, которые не являются частью экспериментальных данных, т.е. используют процесс, называемый ***генерализация***. Таким образом, нейронные сети могут быть использованы для решения реальных мировых проблем как в науке, так и в бизнесе и индустрии [3].

Процесс выбора параметров нейронных сетей, позволяющий приблизительно установить функциональную зависимость между входом x и целью t называют *обучением*. В этом процессе нейронные сети обучают модель работать, используя примеры. Имеется так же арсенал различных методов обучения нейронных сетей, реализация большинства которых требует использование большого количества алгоритмов, что существенно усложняет задание даже при ограниченном количестве циклов. Необходимость в таких итеративных алгоритмах обуславливает высоко нелинейную природу моделей нейронных сетей, что существенно усложняет, а зачастую делает невозможным решение задачи. Итеративный обучающий алгоритм шаг за шагом регулирует вес сигнала в нейронных сетях так, что любое входное значение x в нейронной сети может произвести выходные данные, максимально приближенные к цели t. Поскольку обучающие нейронные сети требуют инициализации итеративного алгоритма при каждой регулировке веса, должны быть заданы начальные условия для приемлемых стартовых величин. Это иногда может потребовать не только качества решений, но и времени для подготовки обучения. Для получения точных данных важно, чтобы инициализация весов начиналась с небольших значений в линейном режиме, которые затем постепенно увеличивались.

Нейронные сети сами по себе не могут прогнозировать событие, до тех пор, пока не пройдут процедуру обучения на примерах. Примеры обычно состоят из пар информации “вход-прогноз”, которые используют одну за другой по мере процесса обучения. Можно представить входные данные в виде “вопроса”, а предсказание в виде “ответа”. Таким образом обучение идет в постоянном режиме вопрос-ответ. Тем не менее, в каждом шаге обучения нейронная сеть требует установку значений весов, используемых данных и оценивает их корректность с помощью критерия, называемого *функцией ошибок*. Если вес выбран недостаточно правильно, производится его корректировка, отвечающая более правильному указанию. Существуют различные технические приемы для решения проблемы правильного выбора и подстановки выходных данных процесса генерализации. Наиболее распространенным является возведение в степень использования данных тестирования.

Использование нейронных сетей для определения технического состояния двигателей внутреннего сгорания (ДВС), в том числе судовых, в настоящее время является перспективным направлением, т.к. нейронные сети обладают несомненными преимуществами по сравнению с другими методами распознавания образов, такими как метод Байеса, дискриминантный анализ, метод ближайших соседей и пр. К основным преимуществам математических моделей основанных на нейронных сетях в задачах классификации можно отнести высокую гибкость и точность указаний, приобретающих большое значение при диагностике технических стохастических систем, к которым можно отнести ДВС.

Моделирование нейронной сети начинается с постановки класификационной задачи, включающей в себя два ответственных этапа: 1 - сбор и анализ информации о наиболее распространенных неисправностях, соответствующих определенной модели силовой установки; 2 – определение состава диагностических признаков. Далее необходим трудоемкий лабораторно-стендовый эксперимент (создание массива выборки) с целью определения влияния искусственно созданных неисправностей на изменение значений диагностических признаков. Трудоемкость экспериментальных исследований может быть существенно снижена за счет использования метода планирования эксперимента, позволяющего моделировать взаимное влияние неисправностей. В дальнейшем проводится разделение массива данных на две или три подгруппы. Первая из которых – обучающая выборка, предназначена для обучения нейронной сети, т.е. анализа зависимости между входными и выходными данными, создания модели нейронной сети, предсказывающей с минимальной ошибкой. Вторая предназначена для

экспертизы качества нейронной сети, используется в режиме тестирования. Третья – утверждающая, используется для окончательного контроля.

Генерализация данных, отвечающая наибольшей точности определения класса, является приоритетом при моделировании нейронной сети. Значительное количество исходных данных (выборки) и, следовательно, наличие помех существенно осложняет задачу классификации.

Автоматизированные нейронные сети ANS (например в пакете *Statistica10)* на начальной стадии моделирования позволяют выбрать процентное соотношение обучающей и контрольной выборок, коды для использования обучающей и контрольной выборок, тип нейронной сети, минимально и максимально допустимое количество нейронов во внутреннем слое. ANS по завершении проверки работы нейронной сети указывает точность при обучении и тестировании, а также за счет каких алгоритмов и функций это достигнуто [4].

Таким образом, построение эффективной нейронной сети сводится к множественному перебору разнообразных параметров сети в поиске оптимального решения. Далее с помощью утверждвющей выборки производится проверка точности прогнозирования нейронной сети. В случае получения удовлетворительнго результата синтезированная нейронная сеть признается пригодной для определения технического состояния объекта.

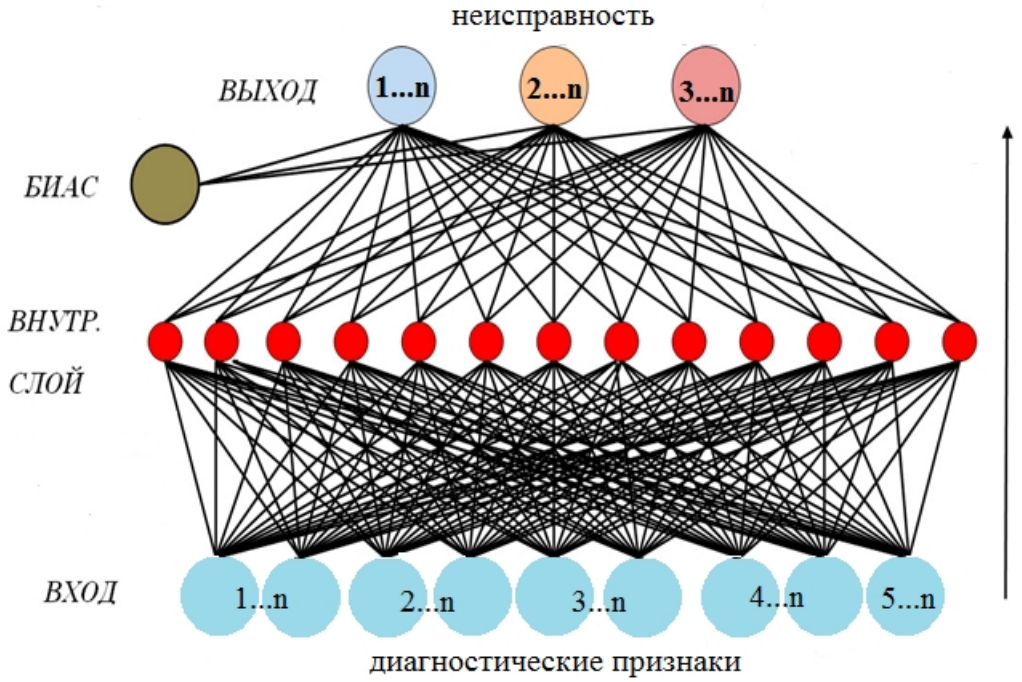


Рисунок 2. Примерная архитектура нейронной сети для диагностики энергетической установки

Нейронные сети по своей архитектуре могут быть как на основе многослойных перцептронов (MLP) так и с использованием радиально- базисной функции (RBF). Как MLP, так и (RBF) содержат в своей структуре биас. Биас это нейрон, в котором активационная функция постоянно имеет единичное значение. Сеть (RBF) в отличие от MLP сети содержит биас только в выходном слое. В большинстве случаев RBF сеть проще чем MLP. BRF так же имеет однонаправленные полные связи между нейронами различных слоев, обеспечивающие работу механизма прогнозирования событий.

Каждый нейрон во входном слое создает весовую сумму донного входного сигнала и передает ее через трансферную функцию для получения значений выхода. Это означает, что в отличии от MPL, RBF сети имеют два типа параметров: 1 - размещение и распространение базисной функции, и 2 - определение веса (значимости) и соединение базисной функции с выходными нейронами.

На рисунке 2 показан пример радиально-базисной нейронной сети предназначенной для самодиагностики судовой энергетической установки.

Входной слой такой нейронной сети представляет собой совокупность диагностических признаков, которая для силовой эергетической установки может состоять из нескольких групп,

например: диагностические признаки первой группы 1-n (давление в цилиндрах; в магистралях высокого и низкого давления топливной системы; разряжение во входном воздушном трубопроводе; расхрд топлива и пр.); диагностические признаки второй группы 2-n (уровень вибрации в области каждого из цилиндров, уровень вибрации в опорах и пр.);диагностические признаки третей группы 3-n (общий уровень шума на различных режимах работы и пр.);диагностические признаки четвертой группы 4-n – тепловые диагностические признаки (температура выпускных газов по цилиндрам, температура охлаждающей жидкости и т.д.); диагностические признаки пятой группы 5-n – электрические диагностические признаки (напряжение на запальных свечах, напряжение на реле подачи топлива, длительность электромагнитного импульса на форсунке и пр.).

Выходной слой представляет собой совокупность возможных неисправностей, которые, в свою очередь, могут разделяться на группы: 1…n– неисправность кривошипно шатунного механизма по цилиндрам, неисправность механизма газораспределения по цилиндрам, механические повреждения узлов двигателя и пр.; 2…n – неисправность топливной системы с указанием конкретного узла, или причины ее вызвавшей и пр.; 3…n – неисправность, вызванная отказом одного или нескольких элементов электрооборудования.

Синтез аналогичных нейронных сетей для автотракторных двигателей показал, что точность определения технического состояния может достигать 99%, однако основные трудности реализации модели самодиагностики заключаются в следующем:

1. Высокая трудоемкость сбора и обработки информации по откзам и неисправностям.
2. Необходимость проведения широкомасштабного лабораторного эксперимента по созданию массива обучающей выборки.
3. Для точного прогнозирования остаточного ресурса агрегатов и двигателя в целом необходимо проведение дорогостоящих ресурсных испытаний.

Заключение

Анализ существующих методов самодиагностики показал, что нейронные сети, управляемые посредством сложных математических функций, могут использоваться для получения образов, распознавания неисправностей судовых энергетических установок и решать задачи такой сложности, которые не под силу для математических моделей, использующих аналитические или параметрические методики. Это обусловлено способностью глубокого анализа значений диагностических параметров и возможностью использования опыта, полученного в результате "обучения” на основе экспериментальных данных, т.е. использования процесса, называемого *генерализацией.* Одним из достоинств нейронных сетей является способность точного определения технического состояния объекта, определения неисправностей и прогнозирования ресурса деталей и узлов судовой силовой энергетической установки. Синтез сетевых моделей систем управления и самодиагностики является перспективным научным направлением, позволяющим повысить надежность и долговечность силовой установки за счет использования искусственного интеллекта при определении технического состояния объекта и дефектов на ранней стадии их проявления.

ЛИТЕРАТУРА

1. Веремеенко А.А., Полуян А.С.Обзор существующих систем управления двигателями// Транспорт и транспортно-технологические системы. Материалы международной научно-технической конференции. Тюмень:Тюменский индустриальный университет,2017. С.87–91

2.Надеев А.И, Буй Хан Нгок, Свирепов Ф.Д. Интеллектуальное управление судовыми двигателями внутреннего сгорания с учетом диагностики состояния оборудования.// Вестник государственного астраханского университета. 2011. №1. С.45–50.

3. Журавлев, Ю.И. Распознавание. Математические методы. Программная система / Ю.И.Журавлев, В.В.Рязанов, О.В.Сенько. – М.: Фазис, 2005. – 159 с

4. Колпаков В.Е. Искусственный интеллект в определении технического состояния диагностируемого объекта / / Известия Санкт-Петербургского государственного аграрного университета. – 2014. – №36. – С. 263-271

V.E. Kolpakov (St. Petersburg Naval Academy)

Outlook of using neural networks for control diagnostics of ship power plant

Modern scientific and technological progress leads to constant complication of technical systems and, as a result, to an increase in labor and material costs associated with the operation of the equipment. Under these conditions, the widespread introduction of diagnostic systems is becoming one of the most important factors to improve the efficiency of power plants.

Despite the common approaches, the self-diagnosis systems of marine power plant engines have some differences from the high-speed and relatively small-sized automotive tractor engines. For example, the self-diagnosis system (CEDS) of the ship diesel engines of the company "Sulzer" (Switzerland) analyzes the current parameters of the diesel and its fuel system. If you change work diagnostic parameters is the analysis of trends in time and set the desired term of the bulkhead site. Failure of the node is warned by the alarm of reaching the limit value of the parameter. However, self-diagnosis system does not currently reached a state in which it would be possible to rely on their information. Thus, the diagnostic system does not cover mechanical damage to the engine or side effects caused by vacuum leakage or malfunction of one or more elements of the gas distribution system, which can cause problems with the composition of the mixture or idling.

Thus, the development of engine management systems and self-diagnosis of the new generation is a complex task, which includes both the use of modern diagnostic tools and the improvement of the mathematical apparatus for determining the state of the diagnosed object. Modern systems of technical diagnostics should be directly related to the recognition of the technical condition of the object, one of the promising areas of development of which is the use of neural networks for self-diagnosis of the power plant.

Neural networks controlled by means of complex mathematical functions can be used to obtain images, to recognize trends and to solve problems of such complexity that are beyond the power of mathematical models using analytical or parametric methods. This is due to the ability of in-depth analysis of the values of diagnostic parameters and the possibility of using the experience obtained as a result of "training" on the basis of experimental data, i.e. the use of a process called generation. One of the advantages of neural networks is the ability to accurately determine the technical condition of the object, fault detection and resource forecasting of parts and components of the ship's power plant.

МЕТОДЫ ЦИФРОВИЗАЦИИ ИНФОРМАЦИОННОГО ПОТОКА В КЛИНИЧЕСКОЙ СФЕРЕ ЗДРАВООХРАНЕНИЯ

В. С. Блюм[1]
(*СПИИРАН, С.-Петербург, 14-я линия, 39, lai_spiiras@iias.spb.su*)

Введение

Цифровизация клинической (лечебной) медицинской информации – процесс объективный и необходимый. Необходимость этого процесса следует связывать, в первую очередь, с неуклонно возрастающим показателем смертности от такого социального заболевания как дефекты оказания медицинской помощи. Если в 1999 году исследования указывали на 10% смертей в госпитальной сфере США, как следствие врачебных ошибок [1], то согласно данным, опубликованным в British Medical Journal в 2016 году, в рейтинге причин смертности, дефекты оказания медицинской помощи заняли уверенное третье место, после сердечно-сосудистых и онкологических заболеваний. Ежегодно от этих причин погибает около 250 тысяч американцев. Как обстоит дело в России, нам ещё только предстоит узнать в результате цифровизации информационной составляющей клиничекой сферы здравоохранения.

Раннее обнаружение дефектов оказания медицинской помощи всегда было и остаётся одной из наиболее актуальных и острых проблем в сфере здравоохранения.

Ожидания качественных изменений в способах решения проблемы раннего обнаружения дефектов оказания медицинской помощи связывают с переходом на электронный документооборот в клинической сфере здравоохранения. Использование компьютера для формирования персональных медицинских записей (ПМЗ) врачами и диагностическими лабораториями, а также интегрирование ПМЗ в электронные медицинские карты в региональных и федеральном центрах обработки данных (ЦОД) создают условия для превращения системы здравоохранения в социальную иммунную систему, которая не только непрерывно контролирует состояние каждого гражданина страны, но и способна оперативно концентрировать необходимые ресурсы для борьбы с вторжениями болезней. Достичь указанный эффект возможно только путём цифровизации полного потока клинической информации.

Основой (ядром) системы охраны здоровья России (рисунок 1) является её клиническая компонента, то есть та часть системы, в которой реализуются лечебно-диагностические процессы (ЛДП). Исполнительными органами ЛДП являются дипломированные медицинские сотрудники и лицензированные медицинские диагностические лаборатории. Именно эти участники ЛДП образуют два класса квалифицированных источников медицинской информации (КИМИ), которые формируют поток первичной медицинской информации в клинической сфере здравоохранения.

Ещё недавно сбор и оперативная обработка подобного потока сложной информации, источники которой распределены по всей стране, были технологически не реализуемы. Напомним, что годовой объём первичной медицинской информации составляет 3–5 терабайта [2], с интенсивностью более миллиона персональных медицинских записей в сутки, при большом структурном разнообразии и высокой семантической сложности.

В настоящее время проблема сбора и обработки «больших данных» – полного потока клинической информации перешла в технологическую плоскость.

Федеральный сервис «Интегрированная электронная медицинская карта»

В состав Единой государственной информационной системы в сфере здравоохранения Российской Федерации входит федеральный сервис интегрированной электронной медицинской карты [http://zakupki.gov.ru], который должен быть реализован с использованием сервисно-ориентированных и облачных технологий. Решение проблемы цифрового представления полного, достоверного и доступного для автоматизированной обработки массива первичной меди-

[1] Кандидат технических наук, старший научный сотрудник.

цинской информации лежит в основе цифровизации здравоохранения. Как указано в техническом задании, «основной целью ведения Интегрированной электронной медицинской карты (далее ИЭМК) является обеспечения непрерывности, преемственности и качества лечения, а также своевременной профилактики и иных мероприятий по обеспечению здоровья конкретного индивида путем документирования и сохранения соответствующей медицинской информации и своевременного предоставления ее уполномоченным медицинским работникам и организациям, а также субъектам ИЭМК.

Под ИЭМК следует понимать коллекцию ЭПМЗ, которая включает все, без исключения, упорядоченные по времени медицинские записи, которые сделали КИМИ по поводу конкретного пациента, независимо от того, в каком ЛПУ или медицинской диагностической лаборатории страны эта запись выполнена.

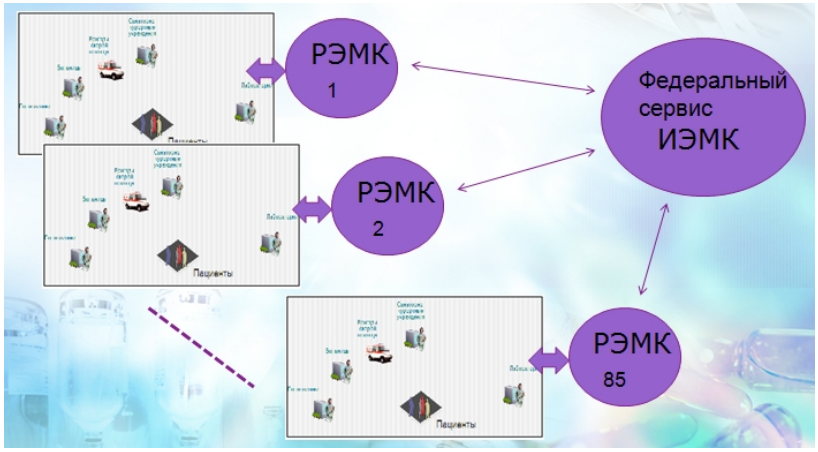


Рис. 1. Информационная инфраструктура клинической сферы здравоохранения России

На рисунке 1 приведена схема трёхуровневой медицинской информационной системы клинической сферы здравоохранения России. Согласно этой схеме поток первичной медицинской информации (поток ЭПМЗ), формируемых КИМИ, имеет три уровня интеграции: в базе ЭМК лечебно-профилактического учреждения (ЛПУ), в базах ИЭМК региональных центров обработки информации (ЦОД), а затем в базе ИЭМК федерального ЦОД. Такой подход гарантирует полноту и достоверность клинической информации для её дальнейшего использования в лечебно-диагностическом процессе.

Электронная история болезни и персональная электронная медицинская запись

Электронная история болезни (ЭИБ) согласно ГОСТ [ГОСТ Р 52636-2006] – это «информационная система, предназначенная для ведения, хранения на электронных носителях, поиска и выдачи по информационным запросам (в том числе и по электронным каналам связи) персональных медицинских записей.» ЭИБ состоит из размещённых в хронологической последовательности электронных персональных медицинских записей (ЭПМЗ). Понятие «электронная персональная медицинская запись» соответствует международному термину EHR – Electronic Health Record [3].

Определяющими свойствами ЭПМЗ являются: во-первых, ЭПМЗ создают квалифицированные источники медицинской информации, во-вторых, ЭПМЗ – это неделимая и неизменяемая запись истории болезни (электронной истории болезни).

Синтаксическая структура ЭПМЗ на метаязыке может быть представлена в виде:

<ЭПМЗ> :: = <время сеанса >< идентификатор КИМИ>
<идентификатор ПАЦИЕНТА>< документ КИМИ> (А.1)
<идентификатор КИМИ> :: = < идентификатор МР > | < идентификатор ДЛ>, где
<идентификатор МР> – идентификатор медицинского работника,
<идентификатор ДЛ> – идентификатор медицинской диагностической лаборатории.

Обратим внимание на два обязательных атрибута ЭПМЗ, на <время сеанса> и < идентификатор ПАЦИЕНТА>.

Обозначим ЭПМЗ, сделанную j-м врачом по поводу i-го пациента, как $z(i,j,t)$, где t – время, указанное в ЭПМЗ, а соответствующее событие обозначим как – $\breve{z}(i,j,t)$. Выделим два особых вида ЭПМЗ и соответствующих событий: записи, в которых врач фиксирует появление (рождение) нового пациента, обозначим – $z_0(i,j,t)$, а записи, в которых врач констатирует смерть пациента, обозначим – $z_f(i,j,t)$. Эти события обозначим как $\breve{z}_0(i,j,t)$, $\breve{z}_f(i,j,t)$.

В случае, когда все ЭПМЗ поступают в ЦОД, можно говорить о том, что интегрированная электронная медицинская карта (ИЭМК) есть функция распределения информации о здоровье пациента. Эта информация добыта системой здравоохранения на всём интервале жизни пациента. Действительно, каждая ЭПМЗ имеет обязательный параметр – время её создания. В результате формирования ИЭМК фиксируется ряд интервалов времени от рождения до смерти пациента, задающих частоту взаимодействий пациента с системой здравоохранения. Кроме того, каждая встреча пациента с КИМИ связана с приращением информации о состоянии здоровья пациента, которая накапливается в ИЭМК. Очевидно, что количество таких приращений соответствует числу временных интервалов. Соответствие временных и информационных интервалов нарушается, только в момент смерти пациента. После этого события информационное приращение (заключение врача о смерти) имеет место, а соответствующий временной интервал отсутствует (рисунок 2).

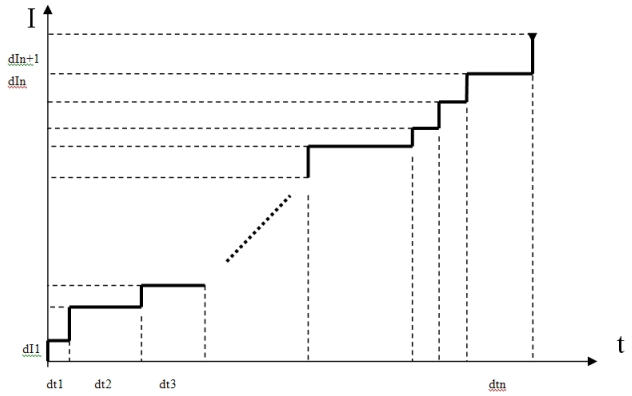


Рис. 2. Функция распределения информации о здоровье в ИЭМК

Рассмотрение информационно-временной плоскости ИЭМК (рисунок 3) показывает, что при увеличении интервала времени между возникновениями двух последовательных ЭПМЗ $\Delta_1 < \Delta_2 < \Delta_3$ уменьшается угол наклона информационного вектора $\alpha_3 < \alpha_2 < \alpha_1$, что может быть использовано в качестве объективного критерия для учёта «оторвавшихся» от диспансерного наблюдения и организации профилактического контроля здоровья с учётом всей предыстории врачебных вмешательств.

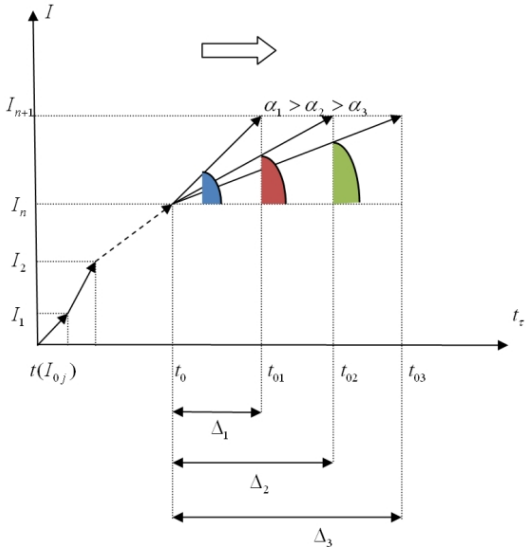


Рис. 3. «Угол внимания» системы здравоохранения к пациенту

В результате анализа информационного вектора ИЭМБ, может быть оценено не только его информационное наполнение (объём ресурсов здравоохранения, затраченных на пациента), но и, так называемый, «угол внимания» системы здравоохранения по отношению конкретному пациенту. Этот параметр важен при организации государственной системы диспансерного контроля состояния здоровья.

Ключевой проблемой формирования и хранения ЭПМЗ является тот факт, что до настоящего времени такой объект, как ЭПМЗ – законодательно не определён. В законе об охране здоровья нет обязательной нормы, которая всякую встречу пациента с врачом или диагностической лабораторией связывала с обязательным возникновением ЭПМЗ. Это обстоятельство существенно снижает доверие к полноте ИЭМК.

Идентификация электронной истории болезни и персональных электронных медицинских записей

Именно врач принимает решение о факте рождении и смерти, поэтому можно утверждать, что система здравоохранения не нуждается в такого рода информации от отделов ЗАГС и других органов государственной власти для организации собственного документооборота. Более того, возникновение в системе новой записи $z_0(i)$, при условии централизованного учёта, может быть использовано для формирования уникального идентификатора истории болезни нового i-го пациента. Для этого достаточно использовать знания о среднем числе рождений в сутки и определить разрядность счётчика: для Петербурга достаточно трёхразрядного числа (около 150 рождений в сутки), для России в целом – четырёхразрядного счётчика (порядка 5 000 рождений), для Всемирной организации здравоохранения (ВОЗ) – семиразрядного числа.

В этом случае, уникальный идентификатор интегрированной электронной медицинской карты будет иметь вид:

<Идентификатор ИЭМК> ::= <дата рождения><учётный номер пациента> (А.2)

Предложенный способ идентификации (2) задаёт отношение строгого порядка на множестве ИЭМК, как в регионе, так и на федеральном уровне. Областью значений грамматической категории <Идентификатор ИЭМК> будет множество нумерованных дат рождения пациентов, а идентификатор i-го пациента – значение функции $I(z_0(i)) = I_{0i}$, для $\forall i$, где I_{0i} – уникальный идентификатор пациента, который одновременно фиксирует возраст пациента относительно текущего момента времени.

Для $\forall i, j \quad I_{0i} > I_{0j}$ или $I_{0i} < I_{0j}$.

В этом случае, конечное множество идентификаторов ИЭМК ($ID = \{I_{01}, I_{02}, ..., I_{0n}\}$) может быть упорядочено на временной оси и интерпретироваться, как множество моментов времени, которые фиксируют события рождения пациентов. При таком подходе текущий момент времени – t_0 и идентификатор ИЭМК определяют реальный возраст пациента. Для заданного множества идентификаторов ИЭМК всегда может быть найдена карта самого возрастного пациента, не содержащая записи $z_f(i)$. Такой идентификатор обозначим как – I_{0s}.

Интервал времени $\Delta = t_0 - I_{0s}$ назовём актуальным интервалом времени жизни популяции. Ось времени, на которой размещены моменты рождения пациентов, назовём осью времени жизни популяции и обозначим – t_ρ. Ясно, что на интервале Δ зафиксированы моменты рождения не только пациентов, требующих текущего внимания системы охраны здоровья, но и тех, в истории болезни которых имеется ЭПМЗ – $z_f(i)$. Такие ЭИБ будем считать закрытыми, но не потерявшими своего информационного значения для системы здравоохранения. Обозначать такие ИЭМК будем как $\widehat{I}_{0i}$. А событие, связанное с появлением ЭПМЗ о рождении нового пациента обозначим – $\breve{I}_{0i}$.

Предложенный способ идентификации ИЭМК и её составляющих – ЭПМЗ, имеет ряд технологических преимуществ по сравнению с любым другим способом.

Психологически пациенту легко запомнить код своей электронной медицинской карты – это всего пять цифр, кроме даты рождения. Факт рождения нового пациента фиксируется именно системой здравоохранения и фиксируется раньше, чем гражданин обретает имя.

Оперативный централизованный учёт нового пациента становится реальной и просто решаемой задачей в связи с переходом на электронный документооборот в здравоохранении и созданием для этих целей федерального центра обработки данных (ЦОД).

Введение в учётную схему нового пациента любого возраста не нарушает этот строгий порядок и технологически не сложен, так как процедура связана с простым добавлением очередной записи в границах учётной даты. Таким образом, предложенный способ идентификации ИЭМК позволит однозначно связать конкретного пациента с каждой, выполненной по поводу его здоровья ЭПМЗ, с его электронной медицинской картой в конкретном лечебно-профилактическом учреждении и, наконец, с его ИЭМК в региональном и федеральном ЦОД.

Для реализации программной поддержки данного способа идентификации пациентов необходима технология идентификации ИЭМК, которая включает ряд программных агентов в федеральном ЦОД, программы идентификации и верификации имени ИЭМК пациента, работающей в каждом ЛПУ и диагностической лаборатории.

Программные агенты, поддерживающие технологию идентификации документов в базе ИЭМК ЦОД, включают:

- Программа контроля входного потока ЭПМЗ,
- Диалоговая программа регистрации новой ИЭМК,
- Программа закрытия ИЭМК.

Процедура создания новой ИЭМК запускается программно по инициативе программного агента ЦОД.

Векторно-событийная модель базы ИЭМК

Для построения событийной модели базы ИЭМК [107] определим ортогональную систему координат, в которой осью абсцисс станет t_ρ – дискретная ось времени жизни популяции, а осью ординат будет непрерывная временная ось, которую назовём осью времени жизни пациентов и обозначим – t_τ. Начало координат определим, как самый ранний момент времени рождения пациента, история болезни которого, по каким-то причинам, сохранилась в архивах системы здравоохранения (рисунок А1), но так, чтобы все наши дальнейшие построения выполнялись в первом квадранте построенной системы координат.

Для самого возрастного пациента, идентификатор ИЭМК которого – I_{0s}, построим вектор $\vec{v}_{0s}$, длина которого $\left|\vec{v}_{0s}\right|$ будет равна реальному времени жизни этого пациента. Вектор $\vec{v}_{0s}$ разместим параллельно оси t_τ с началом в точке I_{0s}. Очевидно, что $\left|\vec{v}_{0s}\right| = \Delta$ – время жизни пациента равное времени жизни популяции.

Покажем, что вектор $\vec{v}_{0s}$ есть проекция информационного вектора истории болезни на временную ось t_τ.

Построим соответствующие вектора $\vec{v}_{0i}$ для всех учтённых ИЭМК (имеющих идентификаторы – I_{0i}), но не имеющих ЭПМЗ – $z_f(i)$. Начало этих векторов будет лежать на оси t_ρ, а конец окажется одной из точек гипотенузы равнобедренного прямоугольного треугольника с катетом $\left|\vec{v}_{0s}\right|$.

Каждый из построенных векторов суть временной образ (проекция) истории болезни конкретного пациента, который, в свою очередь, состоит из суммы векторов, число которых соответствует числу ЭПМЗ в данной ИЭМК.

$\vec{v}_{0i} = \sum_{j=1}^{k} \vec{v}_j$, где k – число записей в ИЭМК i-го пациента, а $|\vec{v}_j|$ – интервал времени от данной ЭПМЗ до следующей, либо до текущего момента t_0. Заметим, что текущий момент времени рассматриваемом в пространстве $T \times T$ характеризуется прямой $y = -t + t_0$

Прямоугольный треугольник, катетами которого являются вектор $\vec{v}_{0s}$ и отрезок оси t_ρ – $[I_{0s}, t_0]$, ограничивает временное пространство $\Omega \subset T \times T$, в котором размещены все, без исключения, актуальные события, которые фиксируют моменты возникновения ЭПМЗ вида (1) в системе здравоохранения (системе ныне здравствующих пациентов). Актуальными будем называть такие события, которые соответствуют ЭПМЗ здравствующих пациентов, то есть пациентов, в ИЭМК которых не содержится записи – $z_f(i)$.

Таким образом, мы построили ограниченное временное пространство Ω, в котором размещены и строго упорядочены ЭПМЗ всех ныне здравствующих потенциальных пациентов (рисунке 4).

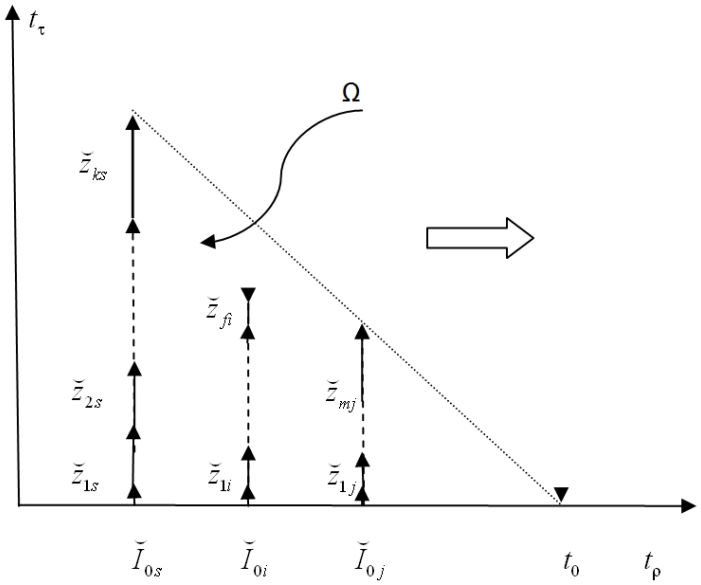


Рис. 4. Модель событий базы ИЭМК

В данной модели для текущего момента t_0 векторные образы ИЭМК всех учтённых живых пациентов упорядочены по величине (времени жизни пациента), а вершины всех векторов лежат на одной прямой, которая является гипотенузой прямоугольного треугольника, ограничивающего временное пространство Ω. В пространстве $\Omega \subset T \times T$ размещены и строго упорядочены по оси t_ρ все ИЭМК, а по оси t_τ строго упорядочены все ЭПМЗ в векторах каждого пациента.

Гипотенуза равностороннего треугольника (пространства Ω) – это отрезок прямой $y = -t + t_0$, длиной $\sqrt{2}|\vec{v}_{0s}|$. Особенностью этого отрезка является то, что он определяет интерфейс системы здравоохранения. Именно на этом ограниченном отрезке, и только на нём, могут возникнуть новые события (ЭПМЗ) в данной модели системы охраны здоровья в текущий момент.

Динамика событий в модели базы ИЭМК

Выбирая шаг h для наблюдения за динамикой развития модели событий базы ИЭМК, очевидно, речь идёт о суточном интервале, необходимо определить алгоритм учёта новых событий и исключения из рассмотрения не актуальных ИЭМК. Понятно, что правила внесения изменений в модель не должны приводить к временным коллизиям.

Пусть $h = t_1 - t_0$. Тогда, прямая $y = -t + t_0$ примет вид $y = -t + t_1 + h$, где $t_0 = t_1 + h$ текущий момент времени (рис.5. Прямые $y = -t + t_0$ и $y = -t + t_1$ задают временное окно, в котором разрешены новые события (возникновение новых ЭПМЗ) в системе здравоохранения. Пространство Ω потенциально увеличивается на величину $\frac{h}{2}(2\Delta + h)$.

Для приведения модели в соответствие с текущим временем t_0 должны быть выполнены следующие операции.

1. Выделение вычислительных ресурсов для ИЭМК новых пациентов, родившихся на интервале h, т.е. формирование новых идентификаторы $I_{0t_1}, I_{0t_2}, ..., I_{0t_n}$ в порядке поступления n заявок (записей $z_{0t_1}, z_{0t_2}, ..., z_{0t_n}$).

2. Для сохранения возможности фиксации точного времени новых событий в группе ИЭМК – $I_{0t_1}, I_{0t_2}, ..., I_{0t_n}$ на интервале h, соответствующие вектора назначаются равными $\vec{h}$.

3. Каждый вектор $\vec{v}_{0i}$, за исключением векторов новорождённых, увеличивается до пересечения с $y = -t + t_0$, причем если на интервале h пациент i не встречался врачом, то $\vec{v}_{0i} + \vec{h} = \vec{v}'_{0i}$ и $|\vec{v}'_{0i}| = |\vec{v}_{0i}| + |\vec{h}|$, но k – число записей в ИЭМК i-го пациента к моменту t_0 не изменяется; если на интервале h пациент i имел одну или более встреч с врачом, то $\vec{h} = \vec{v}_{0i_1} + \vec{v}_{0i_2} + ... + \vec{v}_{0i_m}$, где m – число ЭПМЗ i-го пациента на интервале h и $\vec{v}_{0i} + \vec{h} = \vec{v}'_{0i}$ и $|\vec{v}'_{0i}| = |\vec{v}_{0i}| + |\vec{h}|$, но k – число записей в ЭИБ i-го пациента к моменту t_0 увеличивается на величину m.

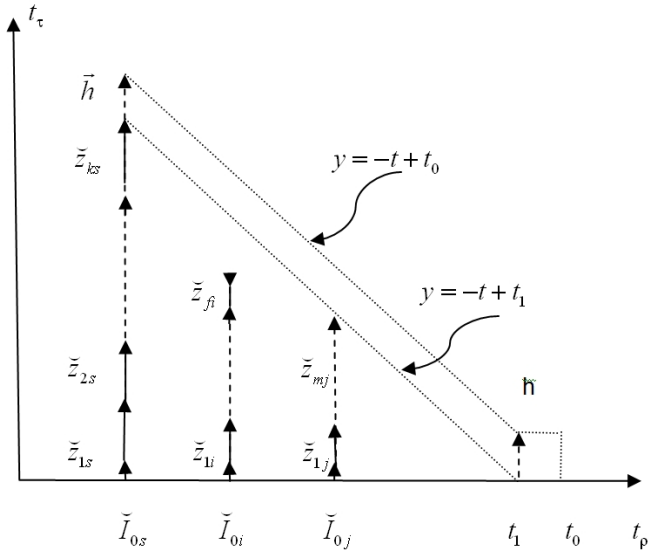


Рис. 5. Динамика событий в модели базы ИЭМК

4. Если на интервале h у i-го пациента возникает запись $z_f(i)$ (вектор – $\vec{v}_{fi}$), то ИЭМК закрывается, обозначим её как $I_{0i} = [I_{0i}]$,а вектор $\vec{v}_{0i}$ обращается в скаляр $|\vec{v}'_{0i}| = |\vec{v}_{0i}| + |\vec{v}_{fi}|$.

5. Если на интервале h у s-го пациента (самого возрастного) возникает запись $z_f(i)$, то выполняется коррекция (уменьшение) пространства Ω. Для этого в списке идентификаторов ИЭМК определяется такой не закрытый I_{0i}, что $\Delta = t_0 - I_{0i} = \max$. Для выбранного идентификатора ИЭМК I_{0i} фиксируется статус I_{0s}.

6. Вычисляются характеристики пространства Ω на момент времени t_0 :

– Число актуальных историй болезни (число живых пациентов);

– Число закрытых ИЭМК на интервале h (число умерших в текущие сутки);

– Коэффициент выживаемости для пространства Ω, а также для любой заданной возрастной категории пациентов (отношение числа актуальных ИЭМК к общему числу учтённых ИЭМК);

– и т.д. и т.п.

Рассмотренный алгоритм указывает на то, что формирование полной, достоверной и доступной для обработки базы ИЭМК предполагает наличие нового класса программных агентов, выполняющих непрерывное обслуживание потока ЭПМЗ в региональных и федеральном ЦОД.

В состав этих агентов с необходимостью входят:

– Программа контроля разрешённого интервала изменений.

– Программа коррекции пространства событий.

– Программа коррекции информационного вектора ИЭМК при отсутствии событий на данном интервале.

– Программа расчёта статистических данных на заданном интервале.

– Программа коррекции интегральных статистических характеристик актуального пространства событий.

– Программа коррекции актуального списка пациентов.

Информационная модель базы ИЭМК

Интегрированная электронная медицинская карта это сумма векторов – ЭПМЗ, которая дополняется системой охраны здоровья на протяжении всей жизни пациента. Начиная с момента рождения, система здравоохранения производит персональные медицинские записи (теперь – ЭПМЗ) и последовательно размещает их в истории болезни (теперь – ИЭМК). Каждая запись имеет собственное информационное наполнение. Не пытаясь на этом этапе оценивать семантику записи, мы всегда можем посчитать количество бит, необходимых для хранения данной ЭПМЗ в памяти компьютера. Ясно, что каждый бит такой записи, в той или иной степени несёт в себе информацию о ресурсах затраченных системой здравоохранения на данного пациента. Очевидно, что каждый вектор (ЭПМЗ) имеет фиксированную по величине (количеству бит) проекцию на информационно-временную плоскость $I \times T_\rho$ (см. рисунок 6).

Особенность предлагаемой информационной модели базы ИЭМК заключается в том, что, при сохранении неизменной величины проекции вектора ЭПМЗ на плоскость $I \times T_\rho$ (в записи не добавляется битов), его проекция на плоскость $I \times T_\tau$ непрерывно увеличивается до момента появления очередной ЭПМЗ, что характеризует процесс старения информации.

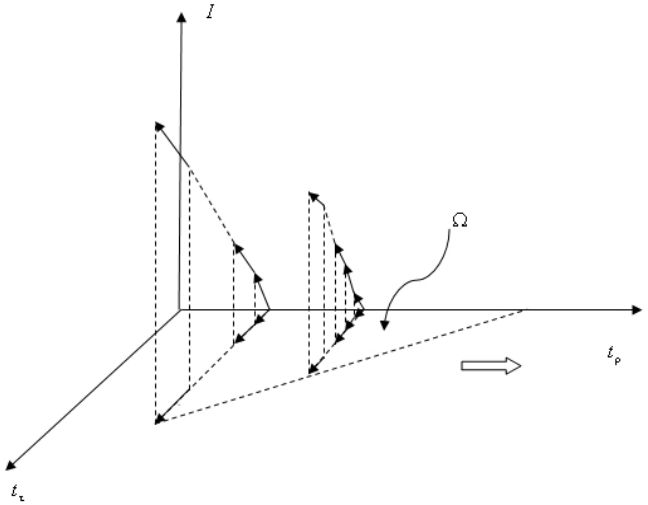


Рис. 6. Информационная модель базы ИЭМК

Вектор последней записи (ЭПМЗ) будет изменять величину и угол наклона к плоскости $I \times T_\tau$ вплоть до появления очередной записи.

Каждая ИЭМК размещена в персональной для данного пациента информационно-временной плоскости ($I \times T$), которая характеризуется моментом рождения – I_{0i} (рисунок 6).

Информационная составляющая ЭПМЗ является объектом семантического анализа и предполагает наличие нового класса программных агентов для оперативной оценки эффективности использования ресурсов здравоохранения и раннего обнаружения дефектов оказания медицинской помощи в результате обработки полной, достоверной и доступной для автоматизированной обработки информации в базе ИЭМК.

Мониторинг состояния цифровой системы здравоохранения по модели базы ИЭМК

Анализ формальной событийной модели базы ИЭМК, которая нацелена на хранение полного, достоверного и доступного для автоматизированной обработки потока первичной медицинской информации, указывает на то, что для функционирования и эффективного использования этой базы данных с необходимостью должны быть реализованы взаимообусловленные коллекции программных агентов, то есть программы, которые выполняют непрерывный мониторинг состояния базы ИЭМК и поддерживают её функционирование.

Важной особенностью предложенной модели являются три эквивалентные способа визуализации текущего момента времени: во-первых, как точки на временной оси, во-вторых, как ограниченной прямой – гипотенуза прямоугольного треугольника (см. рисунок 4) и, в третьих, в виде ограниченной плоскости, проходящей через гипотенузу треугольника событий и перпендикулярной плоскости $\Omega \subset T \times T$ (рисунок 6).

Линейная форма фиксации текущего момента времени позволяет визуализировать всех наблюдаемых пациентов с учетом их возраста.

Фиксация момента времени в виде плоскости позволяет визуализировать всех наблюдаемых пациентов не только с учётом возраста, но и с учётом ресурсов, затраченных системой здравоохранения.

Можно показать, что любого момента времени «плоскость момента времени» (ПМВ) в нашей модели ограничена по оси абсцисс отрезком временной оси равным произведению возраста самого старшего пациента на корень квадратный из двух – Ts *√2, а по оси ординат максимальной величиной информации, содержащейся в одной ИЭМК.

Результат мониторинга состояния пациентов на любом их трёх выше указанных уровней системы охраны здоровья может отображаться точками на ПМВ. Положение точки на плоскости характеризуется двумя параметрами: возрастом пациента и количеством информации в его ИЭМК (объёмом ресурсов затраченных системой здравоохранения на данного пациента) (рисунок 7).

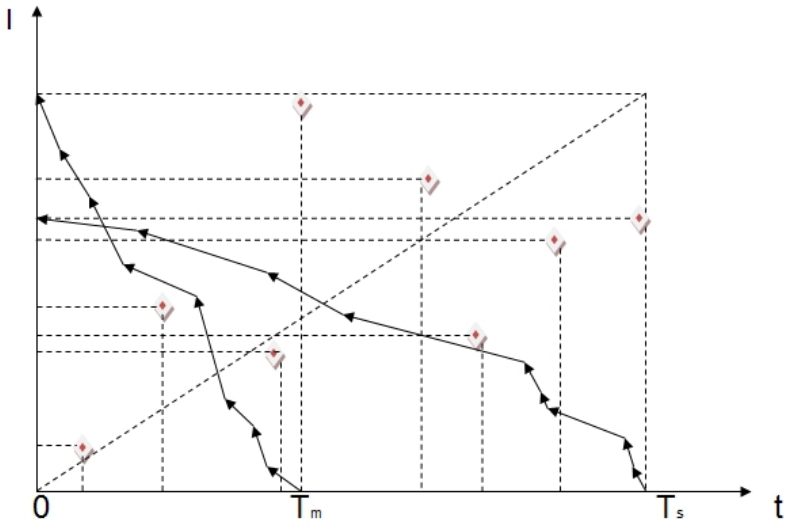


Рис. 7. Мониторинг базы ИЭМК – плоскость момента времени

На рис. 7 представлены точки, которые отражают объём информации а ИЭМК пациентов различных возрастов. Для двух пациентов (самого возрастного и самого «информативного») приведены следы векторных моделей их ИЭМК.

Можно доказать, что плоскости мониторинга различных регионов могут суммироваться, что не приведёт к коллизии данных, так как на множестве ИЭМК задано отношение строгого порядка.

Свойства ПМВ далеко не исследованы, что требует специальных усилий в рамках проекта цифровизации здравоохранения.

Заключение

Разработанная динамическая модель потока первичной медицинской информации представляет полный, достоверный и доступный для автоматизированной обработки набор данных с целью информационной поддержки врачебной деятельности и раннего обнаружения дефектов оказания медицинской помощи. Реализация указанной модели является ключевым методом цифровизации клинической сферы здравоохранения.

Важным шагом на пути внедрения такой модели в практику здравоохранения должна стать фиксация в законе такого объекта как ЭПМЗ и директивное направление потока ЭПМЗ в федеральный ЦОД. Эти цели могут быт достигнуты, например, внесением изменений в ФЗ № 326-ФЗ.

ЛИТЕРАТУРА

1. Kohn LT, Corrigan JM, Donaldson MS. To Err is Human: Building A Safer Health System. Washington, D.C.: National Academies Press, 1999.
2. Блюм В.С., Заболотский В.П. Мысленный эксперимент по организации учета и обработки информационных медицинских услуг. Врач и информационные технологии, №4 2009. стр. 27-35.
3. Международный стандарт. ISO/TR 20514:2005. Информационные технологии в медицине - Электронная медицинская запись - Определение, назначение, содержание (Healthinformatics - Electronichealthrecord - Definition, scopeandcontext).

АНАЛИЗ ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ АППАРАТА КОНЦЕПТУАЛЬНЫХ ГРАФОВ В ЗАДАЧАХ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ МОДЕЛЕЙ ЦИФРОВЫХ ПРОИЗВОДСТВ

С. В. Власенко[1]
(*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина),
С.-Петербург, 197376, ул. Профессора Попова, 5,
s_v_vlasenko@mail.ru*)

Аннотация

Тематика доклада связана с приложением технологий искусственного интеллекта (а именно систем и технологий, основанных на знаниях) в сфере Industry 4.0. Основным предметом исследования является проблематика применения аппарата концептуальных графов при формировании и обработке модельного базиса, используемого в средах платформ «Интернета вещей» (IoT – Internet of Things), ориентированных на поддержку цифровых производств. Непосредственными объектами исследования являются модели знаний различного функционального назначения, построенные на базе концептуальных графов и интегрируемые в общее пространство моделей платформы цифрового производства для решения типовых задач интеллектуальной обработки данных, характерных для выбранной проблемной области. В рамках рассматриваемого подхода указанные модели выступают в качестве структурной основы модельного базиса платформы, обеспечивающей интеграцию разнообразных типов моделей и системных функций на базе единого «интеллектуального ядра».

Ключевые слова: интеллектуальные системы, модели знаний, концептуальные графы, «Интернет вещей» (IoT), цифровые производства

Введение

Развитие технологий Industry 4.0 и «цифровых производств» в настоящее время привело к появлению целого ряда новых комплексных задач, обусловленных спецификой соответствующих технических приложений. К одной из таких задач можно, вероятно, отнести разработку и практическую реализацию подходов к эффективному формированию программных платформ цифрового производства (далее DMP – Digital Manufacturing Platform), обладающих высокой степенью универсальности как в части сфер внедрения и поддерживаемой функциональности, так и в части интеграции в корпоративные аппаратно-программные ландшафты. На сегодняшний день работы в этом направлении ведутся практически всеми ведущими мировыми производителями средств промышленной автоматизации и лидерами смежных сегментов рынка IT-технологий. При этом преобладающей отправной точкой, как правило, является ориентация на развитие выбранных базовых вариантов платформ «Интернета вещей» (Internet of Things – IoT) или «Промышленного интернета вещей» (Industrial IoT – IIoT) за счет «механического» расширения номенклатуры подключаемых функциональных компонентов прикладного уровня, поддерживаемых видов умных устройств, парка API, пользовательских сервисов и т.д. Иными словами, принимаемые за основу механизмы развития платформ не затрагивают принципов их системной организации и функционирования. Между тем, подобные пути наращивания функционального и интеграционного потенциала (и, как следствие, уровня универсальности) DMP исходно обладают рядом естественных системных ограничений, обусловленных, прежде всего, свойствами базовой IoT-платформы и принятой в системе семантической структурой модельного базиса (включая уровень инфологической структуры системного хранилища данных). За пределами данных ограничений модернизация (или перенастройка) имеющихся и внедрение новых компонентов и функций в целевую конфигурацию цифровой платформы становятся весьма сложными и трудоемкими процессами, в том числе – в части проектирования и программной реализации низкоуровневых структурных элементов (близких к системному ядру или входящих в его состав), требующими широкого участия высококвалифицированных специалистов, владеющих необходимым объемом информации о внутренней организации конкретной

[1] Кандидат технических наук, доцент кафедры автоматики и процессов управления СПбГЭТУ «ЛЭТИ».

DMP. В этом отношении среди общих направлений поиска решений рассматриваемой комплексной задачи универсализации DMP особое место занимают подходы, основанные на использовании различных концепций «интеллектуализации» целевых программных платформ и предполагающие определенное видоизменение обобщенных схем их работы. Основной идеей подобных подходов является привнесение в систему способностей «понимать смысл» обслуживаемых производственных процессов и способов воздействия на них (т.е. логику организации системы и взаимодействия ее компонентов при решении целевых функциональных задач), а на основе данного «понимания» – гибко адаптироваться к внутренним трансформациям платформы и обслуживаемой производственной системы, к изменениям во внешнем окружении, а также к профилю решаемых с использованием платформы целевых задач (управление, проектирование, моделирование, анализ, прогнозирование и т.д.). Один из возможных обобщенных подходов описываемого типа предлагается и анализируется в настоящем докладе. Отличительными особенностями рассматриваемого решения могут считаться ориентация на применение специализированных моделей знаний в качестве структурного каркаса единого модельного базиса целевой DMP, а также использование аппарата концептуальных графов (КГ) как базового выразительного средства формального описания данных моделей и процедур их обработки. При этом основное внимание уделяется анализу эффективности применения КГ при решении базовых задач в выбранной проблемной области.

Анализ эффективности применение аппарата концептуальных графов в задачах интеллектуальной обработки моделей цифровых производств

1. Общие проблемы интеллектуализации платформ цифровых производств с применением технологий, основанных на знаниях

Изначально необходимо отметить, что понятия «платформа цифрового производства» (или «предприятия»), как и IoT- или IIoT-платформа на данный момент не могут считаться строго определенными и допускают некоторую свободу интерпретаций. Об этом косвенно свидетельствуют и существенные системные различия решений ведущих компаний, работающих в сфере «Интернета вещей» и цифровых производств, – Microsoft, IBM, Intel, Apple, Amazon, Google, Oracle, Cisco, Siemens, GE, PTC, Rockwell и др. Несмотря на значительное разнообразие предлагаемых продуктов и технологий, представляется возможным выделить несколько определяющих характерных свойств наиболее функционально развитых систем, традиционно относимых к классу «платформ цифрового производства» (DMP) (например, ThingWorx от PTC, Predix от GE, MindSphere от Siemens и т.п.). Среди них наиболее важными, вероятно, можно считать следующие:

- наличие средств поддержки решения задач управления уровней АСУТП и АСУП в реальном времени;
- обеспечение работы с киберфизическими системами и объектами в составе интегрированной информационно-программной среды цифрового предприятия с поддержкой технологий «цифровых двойников»;
- наличие средств AEP (Application Enablement Platform – платформа обеспечения работы приложений) и ADP (Application Development Platform – платформа разработки приложений). При этом под AEP, как правило, подразумевается интерфейсный слой DMP, обеспечивающий возможность работы с хранилищем данных для различных проблемно-ориентированных приложений верхних уровней иерархии управления производством (средства анализа Big Data, включая BI и DataMining, средства машинного обучения, планирования, прогнозирования, бизнес-моделирования, поддержки принятия решений, предиктивного управления ресурсами, визуализации данных и т.д.), а также корпоративных информационных систем различного назначения и уровня интеграции (до уровня ERP включительно);
- использование специализированных экосистем распределенного хранения и обработки данных, на базе которых осуществляются функции интегрированных хранилищ данных, моделей, программных компонентов и т.д. Способы непосредственной реализации указанных функций варьируются в решениях различных разработчиков (облачные/росистые/туманные ресурсы, выделенные сервера, гибридные схемы);

- поддержка технологий IoT на уровне базовых компонентов платформы и «прямых» интерфейсов интегрированного хранилища данных цифрового производства;
- наличие средств, позволяющих адаптировать модельный базис с целью поддержки решения задач слоя «технического проектирования» (инженерного, конструкторского и программного проектирования, математического и имитационного моделирования, подготовки и документационного обеспечения производства и т.п.).

Наглядным примером совокупности всех названных выше свойств можно считать платформу ThingWorx компании PTC, которая, по мнению многих экспертных сообществ (ABI Research, IoT Analytics, Forester Research и др.), является по состоянию на 2018 г. лидером рынка «Умных производств» и PLM-решений, опережая таких конкурентов, как GE Predix, Siemens Mindsphere, Emerson Plantweb, ABB Ability, Bosch IoT Suite, SAP Leonardo, Schneider Electric EcoStruxure, Fujitsu Colmina, Hitachi Lumada, Telit deviceWISE и др.

Дополнительно можно также отметить, что проблемно-ориентированные приложения, интегрируемые в среду цифрового производства посредством AEP-компонентов, в большинстве случаев непосредственно реализуется самим разработчиком DMP и, соответственно, практически входит в состав самой платформы, «скрывая» слой AEP и оставаясь при этом относительно «закрытым» продуктом. К примерам подобного рода можно отнести разработки упомянутой выше компании PTC, интегрируемые с платформой ThingWorx, – широко известную систему Mathcad, CAD-платформу Creo, PLM-систему Windchill и др.

Данное выше описание обсуждаемого класса программных платформ позволяет более точно определить возможное позиционирование «интеллектуальных» компонентов (различного профиля) в общей структуре современных и перспективных DMP.

На основании анализа ряда обзорно-аналитических источников информации ([1-3] и др.) и материалов по продуктам ведущих компаний-разработчиков (GE, Siemens, PTC, Rockwell, Yokogawa, Honeywell, Wonderware, Bosh, Emerson и др.), представляется возможным сделать, помимо прочих, следующие выводы:

- основными сферами приложения концепций искусственного интеллекта в области цифровых производств по-прежнему остаются три основных слоя программных и аппаратно-программных систем и средств: слой «умных» конечных устройств; слой «умных» систем и комплексов средств промышленной автоматизации нижнего уровня иерархии управления технологическими и производственными процессами; проблемно-ориентированные приложения, оперирующие информацией, полученной из единого хранилища данных DMP;
- современные решения класса DMP практически не используют в среде единых хранилищ данных и в общем модельном базисе каких-либо баз знаний или иных компонентов, которые можно было бы отнести к категории моделей знаний. Как следствие, компоненты слоев AEP и ADP не предоставляют пользователям и внешнему программному окружению действительно «интеллектуального» функционала, или реализуют его без использования системных информационных ресурсов платформы;
- функциональность «умных» технических средств, комплексов и систем нижних уровней управления производством (даже в тех случаях, когда эта функциональность в полной мере соответствует формальным признакам поведения интеллектуальной системы, что далеко не всегда справедливо) в подавляющем большинстве случаев проявляется только в среде комплексных решений вполне конкретного и узкого круга компаний-производителей, что обусловлено отсутствием (или недостаточной распространенностью и совершенством) необходимых общепринятых стандартов реализации «интеллектуальных взаимодействий» между системами различных уровней (сказанное относится и к стандартизации в области IoT).

Представленные выше выводы позволяют констатировать, что на текущий момент практическое внедрение концепций построения интеллектуальных цифровых производств носит достаточно ограниченный характер, в том числе – в части применения класса интеллектуальных систем, основанных на знаниях. Между тем, движение в данном направлении могло бы существенно способствовать качественному изменению потенциала предлагаемых решений, включая проблемы наращивания их универсальности, интероперабельности, минимизации затрат на внедрение, эксплуатацию и функциональное развитие и др. Разумеется, разработка подходов к интеллектуализации DMP на базе технологий знаний является весьма обширной областью ис-

следований. Здесь могут рассматриваться и подходы на базе интеграции внешних интеллектуальных систем через слой AEP, и подходы, связанных с переориентацией средств ADP на работу с моделями знаний и методами их обработки, и целый ряд других. В рамках настоящей работы рассматривается только один из возможных альтернативных путей внедрения технологий знаний в системы класса DMP.

Общий смысл данного подхода сводится к формированию целостного, полностью интегрированного в среду платформы интеллектуального ядра, которое выполняло бы, по крайней мере, следующие функции:

- поддержка построения и обработки моделей знаний различного назначения, объединяемых в единое целое на основе комплексной метамодели знаний DMP. К основным типам указанных моделей знаний следует в данном случае отнести:

- модели знаний о предметной области конкретного производства, включая знания о технологических и производственных процессах;
- модели знаний о модельном базисе, используемом в DMP и интегрированной с нею целевой среде;
- модели знаний о структуре систем и средств автоматизации (включая киберфизические подсистемы);
- модели знаний об информационных потоках и ресурсах;
- модели знаний о структуре, составе и функциях компонентов программного обеспечения различных уровней;
- модели знаний о функциональных задачах и способах их решения;
- модели знаний о самой платформе и ее конфигурации;
- прочие типы частных моделей знаний;

- поддержка логической структуризации единого информационного хранилища платформы и интеграции всех компонентов модельного базиса на основе общей метамодели знаний DMP и частных моделей знаний;
- поддержка конструирования, обслуживания и исполнения абстрактных моделей знаний процессного и процедурного типа (т.е., паттернов процессов, обобщенных алгоритмов, имитационных моделей и т.п.), включая непосредственную реализацию решений функциональных задач DMP, представленных в виде моделей соответствующего вида;
- поддержка взаимодействия с внешними интеллектуальными системами, оперирующими аналогичными (или приводимыми) классами моделей знаний на уровне ядра DMP;
- поддержка интеграции с произвольными внешними информационными системами с использованием моделей знаний, описывающих функциональные, информационные, программные и др. виды межсистемных взаимодействий в терминах моделей DMP;
- реализация функций «посредника» между ядром DMP и слоями AEP и ADP;
- поддержка функций слоя ADP в задачах интеллектуального проектирования и моделирования (включая построение компонентных моделей по принципу «plug and play», использование готовых прототипов и шаблонов проектирования, автоматического формирования программного окружения моделей и т.п.).

Одним из первичных вопросов, встающих при реализации интеллектуального ядра DMP, поддерживающего перечисленные выше функции, является, безусловно, выбор наиболее эффективного класса моделей знаний и соответствующего формального аппарата для представления знаний и методов их обработки. При этом с учетом специфики рассматриваемой проблемной области к указанному выбору могут быть предъявлены следующие приоритетные требования:

- использование классов моделей знаний и методов их обработки, хорошо согласующихся с профилем функциональных задач, решаемых на базе DMP;
- эффективность процессов обработки с учетом требований работы DMP в реальном времени;
- необходимость поддержки совместного использования декларативных, процедурных и гибридных моделей знаний в рамках единого механизма их формального описания;
- способность к представлению частных моделей знаний различных типов, перечислявшихся ранее, и к эффективной интеграции этих моделей на базе единой метамодели знаний;

- наличие стандартизированных выразительных средств (языков), которые могли бы использоваться в качестве базовой формы представления моделей знаний выбранных классов при разработке и эксплуатации DMP;
- ориентация на технологии обработки знаний, поддержанные необходимым парком инструментальных средств, и исходно ориентированные на работу с распределенными вычислительными ресурсами.

Среди потенциально возможных подходов, в значительной степени отвечающих рассмотренным выше требованиям, вероятно, следует выделить подход на базе применения онтологических моделей. Онтологии на данный момент являются достаточно хорошо проработанным теоретическим базисом, поддержаны в рамках глобального проекта Semantic Web консорциума W3C, имеют общепринятые «стандарты» описания (в виде рекомендаций W3C по языкам RDF/RDFS, OWL и языку запросов SPARQL) и широко используемый стек технологий программной поддержки. Тем не менее, применительно к исследуемой проблеме формирования интеллектуального ядра DMP следует указать и на ряд недостатков онтологического подхода. Наиболее важными среди них в данном случае являются следующие:

- ограниченная выразительная мощность (в части механизмов образования типов и агрегирования типов и объектов, описания и типизации отношений произвольного вида, включая ассоциативные связи различных разновидностей, отображения полиморфизма типов и объектов, представления контекстных зависимостей, гибкой идентификации субъектов модели и др.);
- отсутствие средств прямого отображения процедурных моделей знаний;
- сложность логической структуризации компонентов модели;
- потеря гарантий поддержки дескрипционных логик при выходе за пределы диалекта OWL 2 DL (т.е., в наиболее выразительном диалекте OWL 2 Full), а также неоднозначность оптимального выбора профилей (EL, QL, RL) в рамках DL.

В значительной мере указанных недостатков лишен рассматриваемый далее альтернативный подход, основанный на использовании теории концептуальных графов (КГ).

2. Применение аппарата концептуальных графов

КГ как класс моделей знаний ([4,5]) обладают рядом отличительных особенностей, которые хорошо согласуются с определенными в разделе 1 требованиями. К достоинствам аппарата КГ, наиболее важным в выбранной проблемной области, относятся, прежде всего, следующие:

- высокая степень универсальности, позволяющая представлять на основе КГ практически любые наиболее распространенные базовые классы моделей знаний и данных;
- способность КГ к представлению как декларативных, так и процедурных знаний на уровне базовых нотаций и структур знаний;
- наличие стандартизированных (ISO/IEC 14481:1998 и ISO/IEC 24707:2007) средств описания (языки CGIF и CLIF, XML-производный язык XCL, графическая нотация DF);
- наличие аппарата формальной логической интерпретации, определенной в так называемой «общей логике» (Common Logic - CL, ISO/IEC 24707:2007), что позволяет строить математически корректные машины «исполнения» КГ. Уместно также заметить, что все языки, регламентированные в рамках Semantic Web, являются семантическими подмножествами языков CL;
- гибкая структуризация КГ (включая рекурсивные определения КГ), обусловленная спецификой базовых видов сущностей моделей – «концептов» и «отношений»: узлы-концепты могут представлять собой самостоятельные КГ, что позволяет формировать произвольные агрегаты из субъектов модели, а также композиции отдельных моделей, причем сказанное относится и к возможностям образования типов объектов и концептуальных отношений в графе (т.к. пространство типов также представляется в виде КГ, позволяющих задавать произвольные семантические связи между типами);
- поддержка непосредственного отображения в КГ контекстно-зависимых и ассоциативных связей между субъектами модели (или их типами), что позволяет гибко оперировать структурными элементами моделей при обработке единого модельного базиса;
- поддержка описания параметризуемых типов концептов и отношений на базе механизма λ-выражений;

- наличие специализированного механизма «неопределенных» (в текущем контексте) отношений («акторов»), которые могут представлять собой, в том числе, вычислительные процедуры произвольного содержания с заданными в КГ входами и выходами;
- наличие механизма «кореферентных ссылок», позволяющего непосредственно устанавливать соответствия элементов различных терминологических базисов, пространств идентификаторов и структур данных, разнородных частных моделей, используемых в среде DMP и т.п.;
- наличие регламентированного (в нотации CGIF) унифицированного способа представления баз знаний (за счет использования предопределенных специализированных контекстов);
- наличие достаточно широкого спектра инструментальных средств различного назначения для работы с аппаратом КГ (Amine, CharGer, Corese, CPE, Cogitant, Notio, CARE и др.) ([6]).

Перечисленные свойства КГ позволяют рассматривать данный аппарат как возможное средство реализации интеллектуального ядра DMP. При этом в обобщенной структуре ядра предполагается наличие, как минимум, следующих компонентов, ориентированных на работу с моделями знаний:

1) виртуальная машина КГ, в функции которой входит непосредственная обработка моделей знаний и управление исполнением процедурных сценариев, представленных в виде КГ, обслуживание трансформаций единого модельного базиса DMP и координация функций системы управления хранилищем данных, а также обслуживание запросов внешних систем на предоставление доступа к данным и программным компонентам платформы, включая запросы слоев AEP и ADP;

2) интегрированная среда разработчика (являющаяся автономным или встроенным компонентом слоя ADP), поддерживающая функции конструирования (с использованием как текстовых, так и графических КГ-нотаций), обслуживания, визуализации и мониторинга исполнения компонентов модельного базиса DMP, построенных или интегрированных на базе КГ;

3) интегрированный модуль трансляции КГ-моделей, реализующий приведение исходного представления КГ к требуемому (в рамках DMP) виду и размещение соответствующей модели в едином хранилище данных платформы (при координации виртуальной машины КГ);

4) интерфейсные модули интеллектуального ядра DMP, поддерживающие его взаимодействие с хранилищем данных и слоями AEP/ADP, а также (при необходимости) прямые взаимодействия с внешними приложениями.

С точки зрения организации общей схемы функционирования DMP указанные компоненты в совокупности обеспечивают исполнение интеллектуальным ядром функций высокоуровневого посредника во взаимодействиях между разнородными программными средствами и информационными ресурсами, привлекаемыми к решению конкретной функциональной задачи. При этом все участники и сам процесс взаимодействия, а также процесс решения задачи описываются КГ-моделями, интегрируемыми на основе метамодели знаний DMP и исполняемыми под управлением виртуальной машины КГ.

Описанный общий подход, основанный на применении КГ, позволяет (помимо обеспечения прочих возможностей, исходно заложенных в идеологию интеллектуализации DMP):

- существенно повысить универсальность и масштабируемость платформы, а также ее способность к наращиванию функционального и интеграционного потенциала (за счет поддержки введения в целевую систему новых компонентов путем формирования соответствующих частных моделей знаний, описываемых в принятой разработчиком терминологической системе, без существенных трудозатрат на перепрограммирование низкоуровневых программных компонентов);
- унифицировать процессы решения произвольных функциональных задач на основе единого формального аппарата их описания;
- обеспечить расширение спектра решаемых на базе DMP функциональных задач за счет построения разнообразных, гибко настраиваемых сценариев решения, представляемых в виде соответствующих КГ-моделей и допускающих комбинирование работы различных средств платформы и интегрированных с нею проблемно-ориентированных систем;
- обеспечить логическую интеграцию различных по своей природе системных ресурсов (информационных, программных, аппаратных) на базе единой метамодели знаний;

- реализовать строго формальный механизм контроля логической и информационной корректности внутрисистемных преобразований;

- обеспечить необходимые условия для эволюционного расширения парка поддерживаемых в среде DMP типов и моделей оборудования, вычислительных средств, программных систем и т.д. на основе «понятийной» систематизации информации.

Приведенный выше перечень (который, вероятно, может быть расширен при более детальном рассмотрении свойств интеллектуального ядра DMP) в целом, возможно, позволяет говорить о внесении в интеллектуальные платформы (построенные по рассматриваемой архитектурной схеме) принципиально новых возможностей в сравнении с традиционными способами их организации. Однако на пути реализации предлагаемых системных решений стоит и целый ряд препятствий как прикладного, так и теоретического плана. К числу наиболее существенных проблем подобного рода представляется необходимым отнести следующие:

1) стандарты CL (особенно в части нотаций CGIF) в их текущем состоянии, возможно, требуют определенной доработки и расширения с учетом требований однозначности интерпретации КГ применительно к задаче реализации универсальной виртуальной машины КГ. Кроме того, недостаточно детализированными в стандартах ISO/IEC представляются и некоторые математические аспекты логических счислений на КГ;

2) недостаточно исследованной на данный момент является проблема оценки эффективности процессов обработки КГ в «мягком» реальном времени, а также в распределенных и параллельных вычислительных средах;

3) весьма неоднозначной в теоретическом плане остается задача поиска оптимальных форм вычислительной реализации сетевых моделей рассматриваемого класса, особенно применительно к необходимым в рамках аппарата КГ классам алгоритмов их обработки;

4) в настоящее время представленные на рынке инструментальные средства, ориентированные на работу с аппаратом КГ, не в полной мере отвечают требованиям действующих стандартов CL и/или не обеспечивают в требуемом объеме и качестве поддержку целого ряда функций и сервисов первоочередной (для разработчиков) необходимости (интерактивное конструирование и визуализация, управление хранилищем знаний и данных и т.д.);

5) реализация DMP на базе аппарата КГ при масштабных внедрениях потребует наличия соответствующих кадровых ресурсов инженеров-системотехников, владеющих необходимыми знаниями в области теории КГ и инженерии знаний в целом, что на текущий момент представляется достаточно проблематичным.

Тем не менее, следует отметить, что перечисленные выше проблемы в целом не представляются неразрешимыми, в связи с чем развитие предлагаемого подхода к построению интеллектуальных DMP, вероятно, можно считать вполне реализуемым.

Заключение

В настоящее время на кафедре автоматики и процессов управления СПбГЭТУ «ЛЭТИ» ведутся инициативные работы в направлении создания варианта интеллектуальной платформы цифрового производства, базирующейся на использовании аппарата КГ и соответствующей основным концепциям, изложенным в настоящей статье. Текущая фаза работ по данному проекту связана с частичной «ревизией» ряда положений теории КГ, о которых говорилось ранее, и с разработкой специализированных расширений нотации CGIF, ориентированных на адаптацию аппарата КГ к рассматриваемой предметной области.

ЛИТЕРАТУРА

1. Pethuru Raj, Anupama C. Raman. The Internet of Things: Enabling Technologies, Platforms, and Use Cases. Auerbach Publications, Boca Raton, FL, USA, 2017. 364 p.
2. Al Fuqaha A., Guizani M., Mohammadi M., Aldehari M., Ayyash M., "Internet of Things: A Survey on Enabling Technologies, Protocols and Applications, IEEE Communications Surveys & Tutorials, Volume 17, Issue 4, 2015.
3. Foundational Elements of an IoT Solution by Joe Biron and Jonathan Follett Copyright. 2016 O'Reilly Media. ISBN: 9781492042655.

4. John F. Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co., Pacific Grove, CA USA, 2000. 594 p.
5. Michel Chein, Marie-Laure Mugnier: Graph-based Knowledge Representation - Computational Foundations of Conceptual Graphs. Advanced Information and Knowledge Processing, Springer 2009, ISBN 978-1-84800-285-2, pp. 1-427.
6. Использование аппарата концептуальных графов в задачах представления и обработки знаний: учеб. пособие / С.В. Власенко, Д.А. Варчев; Санкт-Петербургский государственный электротехнический университет им. В.И. Ульянова (Ленина) "ЛЭТИ". - СПб.: Изд-во СПбГЭТУ "ЛЭТИ", 2009. - 64 с.

S.V. Vlasenko (Saint Petersburg Electrotechnical University «LETI», Saint Petersburg)

Analysis of efficiency of conceptual graphs' theory application in tasks (on issues) concerning intelligent processing of digital manufactures' models

The issue of this report is related to the application of artificial intelligence technologies, (namely knowledge based systems and technologies) in the field of Industry 4.0. The main subject of the research is the problem of conceptual graphs' technique usage in forming and processing of model basis applied in environments of the platforms of Internet of Things (IoT), oriented to support digital manufacturing. The immediate objects of the research are knowledge models (with various functional purposes) built on the base of conceptual graphs and integrated into common model space of digital manufacturing platform for the solution of typical tasks of intelligent data processing, which are typical for the chosen problem area. Within the discussed approach the above mentioned models act out as a structural foundation of platform model basis, providing integration of various model types and system functions on the base of a unified «intelligent core».

КОНТРАКТНЫЕ СЕТЕВЫЕ ПРОТОКОЛЫ ВЗАИМОДЕЙСТВИЯ КОГНИТИВНЫХ АГЕНТОВ В РАСПРЕДЕЛЕННЫХ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ УПРАВЛЕНИЯ.

Д. Г. Арсеньев[1]
(*Санкт-Петербургский политехнический университет Петра Великого, С.-Петербург, Гражданский пр., дом 28, vicerector.int@spbstu.ru*)

Д. Е. Баскаков[2]
(*Санкт-Петербургский политехнический университет Петра Великого, С.-Петербург, ул. Политехническая, дом 21, baskakovdmitrii@gmail.com*)

В. П. Шкодырев[3]
(*Санкт-Петербургский политехнический университет Петра Великого, С.-Петербург, ул. Политехническая, дом 21, shkodyrev@spbstu.ru*)

Аннотация

Развивается концепция консенсусной динамики в управлении структурно-сложными производственными процессами на основе контрактных сетевых протоколов взаимодействия в среде когнитивных агентов. Демонстрируется неэффективность классической теории управления сложным организационными структурами в силу их неопределенности, многозадачности и многомерности. В качестве перспективной модели управления рассматривается интеллектуальная мультиагентная система управления.

Ключевые слова: интеллектуальная мультиагентная система управления, когнитивные агенты, консенсусная динамика, сетевые протоколы взаимодействия.

Введение

Современная инфраструктура промышленных предприятий и производственных процессов отличается крайне высокой сложностью взаимодействующих систем. Иерархические информационные потоки в таких системах управления носят распределенный характер, что приводит часто к возникновению нештатных ситуаций. Традиционные парадигмы управления отличаются низкой эффективностью в условиях сложных сетевых структур с большим количеством взаимодействующих распределенных узлов и параллельными информационными потоками. Это и обуславливает необходимость создания новых принципов управления подобными объектами на основе использования моделей консенсусной динамики с применением интеллектуальных когнитивных агентов. Предлагаемый доклад и посвящен современным парадигмам интеллектуального управления на основе контрактных сетевых протоколов взаимодействия когнитивных агентов в распределенных интеллектуальных системах управления.

Контрактные сетевые протоколы взаимодействия когнитивных агентов в распределенных интеллектуальных системах управления

Наиболее перспективным подходом к управлению сложными производственными структурами является использование интеллектуальной мультиагентной парадигмы. Следует заметить, что понятие интеллектуальной системы управления занимает центральное место в современной теории управления. Под интеллектуальной системой управления мы понимаем прежде всего систему управления, которая способна выделять, аккумулировать и упорядочивать знания об

[1] Доктор технических наук, профессор, проректор по международной деятельности Санкт-Петербургского политехнического университета Петра Великого.
[2] Аспирант Высшей школы киберфизических систем и управления Санкт-Петербургского политехнического университета Петра Великого.
[3] Доктор технических наук, профессор, директор Высшей школы киберфизических систем и управления Санкт-Петербургского политехнического университета Петра Великого.

объекте или объектах управления, а также самообучаться в самом широком смысле на основе полученных знаний, менять свои целевые функции и алгоритмы управления.

Мультиагентные системы управления являются разновидностью агентного похода к управлению распределенными системами. Агент- некая математическая или программная сущность со своей функцией управления [1]. Агент является частью сложной мультиагентной системы управления. Интеллектуальный агент является прежде всего программной реализацией и наделен функциями искусственного интеллекта.

Интеллектуальная мультиагентная система управления состоит из множества интеллектуальных агентов в структурно-сложной иерархии системы управления. Каждый агент обладает следующими свойствами:

1. Автономность – способность действовать в соответствии со своей целевой функцией под влиянием внешнего воздействия и способность контролировать свои действия.
2. Социальность – способность прежде всего обмениваться информацией, извлекать опыт и знания, влиять на модель управления, находить консенсус в случае взаимодействия с иными агентами.
3. Реактивность – способность реагировать на любые внешние воздействия как от других агентов и контуров управления, так и на воздействия окружающей среды (пожар, отключение электричества, например).
4. Проактивность – способность самообучаться на основе полученных ранее знаний, предсказывать или предвосхищать возможные изменения, самообучаться.

Сложность объектов управления приводит и к тому, что исходная информация о них часто бывает неполной, неточной. Весьма часто такие системы следует рассматривать как открытые, то есть активно взаимодействующие с внешней средой. В классической теории мультиагентных систем управления агент не наделен функцией интеллектуальности, то есть способностью к проактивности и самообучению, например.

Следует заметить, что на рынке есть несколько решений для создания универсальных мультиагентных систем управления. Рассмотрим кратко основные из них.

Платформа **JADE (*Java Agent Development Framework)*.** Среда разработки поддерживает стандарт **FIPA (*The Foundation for Intelligent, Physical Agents).*** Данная среду предполагает кроссплатформенную разработку начального уровня с гибким графическим интерфейсом с использованием языка Java не ниже 5.0. Агенты **JADA** обмениваются меду собой с помощью сообщений. Сообщения формируются в различные очереди, что позволяет менять конфигурацию системы. Ключевым недостатком данного продукта является его простота, что не позволяет разрабатывать системы промышленного уровня, тем более интеллектуальные мультиагентные системы. Следует заметить, что последний дистрибутив системы датируется 8 июнем 2017 года, а сам стандарт **FIPA** последней спецификации датирован 2007 годом, что точно не является свидетельством высокой актуальности предлагаемых решений с помощью данной платформы.

Agent Builder является интегрированной средой быстрой и гибкой разработки интеллектуальных систем управления с использованием агентов, а также агент-ориентированных приложений. Разработчик утверждает о полной поддержке всех версий Java, а также языка **KQML (*Knowledge Query and manipulation Language)*** со спецификацией **CORBA (*Common Request Broker Architecture).*** Поддержка разработчика остановилась на версии Windows XP, что говорит об отсутствии какой-либо перспективы данного проекта в его текущем состоянии.

В рамках данной концепции интеллектуального мультиагентного подхода к управлению предлагается рассмотреть следующую парадигму, которая существенно отличается от действующих систем на рынке.

Традиционные агенты управления являются по сути событийными информационными участниками, которые способны участвовать в управлении только на основе получения и передачи той или иной информации, а также действия в соответствии с заданной целевой функцией. Интеллектуальность таких систем существенно ограничена заданной моделью управления, в рамках которой в принципе некорректно использовать данный термин в полном значении. При этом большинство авторов совершенно упускают из виду именно процесс обучения и накопления знаний, а также непосредственно алгоритмов этого обучения.

Именно в рамках создания перспективных моделей и систем интеллектуального мультиагентного управления с использованием когнитивных агентов авторами предлагаются следующие важные концептуальные дополнения, которые качественно меняют существующие парадигмы интеллектуального управления, а именно:

1. Использование теории и математического аппарата консенсусной динамики.
2. Использование теории графов, а также концепции прямого ациклического графа (DAG, *Directed Acyclic Graph)*.
3. Использование в качестве базы знаний графовых баз данных (например, **GraphDB Lite**, **Neo4j Community Edition)**, а также системы машинного обучения типа **H2O**, **TensorFlow** и **Scikit-Learn** [2].
4. Консенсусные алгоритмы взаимодействия между интеллектуальными агентами.

Как ранее сообщалось, агент в нашем понимании- некая математическая сущность, имеющая обычно программную реализацию. Интеллектуальный агент действует рационально и оптимально на основе полученного знания в процессе своего развития и обучения. В сложной распределенной системе управления мы имеем иерархии интеллектуальных агентов, каждый из которых участвует в управлении отдельными подсистемами. Важнейшей особенностью интеллектуального агента в мультиагентной системе управления является наличие помимо базовых целевых функций и моделей взаимодействия с иными агентами использование функций искусственного интеллекта, таких как:

1. Обучение с использованием алгоритмов машинного обучения с учителем (контролируемое обучение): дерево принятия решений, наивная байесовская классификация, метод опорных векторов, метод ансамблей.
2. Обучение с использованием алгоритмов машинного обучения без учителя (неконтролируемое обучение): алгоритмы кластеризации, метод главных компонент, анализ независимых компонент.
3. Нейронные сети.
4. Консенсусные алгоритмы на основе искусственного интеллекта.

При этом важно понимать, что интеллектуальный агент управления, являясь частью сложной мультиагентной системы управления, может иметь различные состояния своих целевых показателей и функций управления, как и вся система в целом. Дополнительно сама система управления должна иметь возможность изменять на ходу как иерархии управления в зависимости от внешних и внутренних воздействий, а также нештатных ситуаций, так и сами алгоритмы управления. Все это требует использование в реальном времени готовых моделей управления, а также иных моделей управления, полученных в процессе обучения, особенно в процессе обучения без учителя.

Для хранения данных моделей управления, а также их возможной модернизации и предполагается использовать именно графовые базы данных. Последние себя зарекомендовали крайне эффективным средством работы с неполными данными, сложно структурированными данными, а также данными, которые имеют иерархическую структуру.

Под контрактным сетевым протоколом взаимодействия мы понимаем прежде всего совокупность алгоритмов и правил взаимодействия интеллектуальных агентов между собой в иерархии распределенной мультиагентной системы управления. Поскольку данные алгоритмы взаимодействия включают в себя и элементы искусственного интеллекта, то считаем возможным в данном случае говорить о когнитивных интеллектуальных агентах управления.

Программная реализация таких систем управления представляет собой нетривиальную задачу, чем и обусловлено полное фактическое отсутствие на рынке неких открытых решений, которые хотя бы частично удовлетворяли поставленной задачи в такой формулировке.

Когнитивные мультиагентные системы управления обладают следующими важнейшими характеристиками, которые существенно влияют на особенности программной реализации данных комплексов, а именно:

1. Высокая конкурентность взаимодействия когнитивных агентов между собой, чем и обусловлено использование методов консенсусной динамики.

2. Крайне высокая степень параллелизма протекающих управляющих и информационных процессов.
3. Большое количество контуров с обратными связями.
4. Обработка информации в режиме реального времени.

Нам представляется крайне перспективным использование парадигмы реактивного программирования для реализации подобных задач. Поскольку такие системы дополнительно насыщены большим количеством внутренних и внешних событий и вызовов, то данные системы управления относятся к высоконагруженным данными приложениям (DIA, *Data-Intensive Applications)* [3]. Мы понимаем под приложением, высоконагруженным данными, такое приложение, в которых качество данных, степень их сложности, а также скорость их изменений являются ключевыми в отличие от высоконагруженных вычислениями приложений (CIA, *Compute*-Intensive Applications).

Таким образом, концепция именно реактивного программирование становится ключевой при разработке подобных мультиагентных систем управления [4].

Важно заметить, что реактивное программирование является частью парадигмы функционального программирования, которая должна являться основой, по мнению авторов, для разработки подобных систем управления.

Наиболее интересной и полезной в процессе написания кода нам представляется использование библиотеки RxJava (Reactive X) совместно с JavaStream [5]. Данная библиотека Java обладает следующими важнейшими преимуществами, которые присущи функциональному программированию:

1. Проталкивание и вытягивание событий.
2. Синхронный и асинхронный режимы работы системы управления.
3. Конкурентность и параллелизм.
4. Ленивые и энергичные типа.
5. Двойственность.
6. Использование реактивных расширений (подписка на уведомления, прослушка уведомлений, горячие и холодные объекты и события).
7. Многопоточность.
8. Периодическая выборка и отбрасывание событий.
9. Противодавление и сопротивление событиям.

Одним из наиболее перспективных направлений использования когнитивных агентов в распределенных интеллектуальных системах управления является область управления цепочками поставок (SHM, *Supply Chain Management)* и логистикой в частности с дополнительным использование концепции IoT (*Internet of Things)* или II (*Industrial Internet).* Использование последних потребует с высокой степенью вероятности внедрения технологий распределенного реестра или blockchain- технологий, что позволяет говорить о создании на основе когнитивных агентов систем управления последнего поколения с расширенными функциями искусственного интеллекта.

Использование интеллектуальных мультиагентных систему управления на основе концепции контрактных сетевых протоколов показывает свою высокую эффективность в части снижения, например, затрат на человеческий капитал. В ходе тестовых испытаний при управлении цепочкам поставок удавалось достичь снижения таких постоянных издержек более чем в 2 раза. При этом общая эффективность работы системы управления оказывалась существенно выше, чем использование традиционных логистических моделей на основе линейного программирования.

Заключение

В докладе рассмотрена концепция контрактных сетевых протоколов взаимодействия когнитивных агентов в распределенных интеллектуальных системах управления. Предложены перспективные направления реализации данных моделей с использованием современных средств разработки. Данные методики и парадигмы коллектив авторов использует в практической работе и создании программных и аппаратных прототипов для отраслевых заказчиков из сферы нефтепереработки и нефтедобычи.

ЛИТЕРАТУРА

1. Д.Г. Арсеньев, Иванов В.М., Шкодырев В.П. Адаптивное управление вычислительным процессами. СПб.: Издательско-полиграфическая ассоциация университетов России, 2016, 315 с.
2. Robinson I, Webber J, Eifrem E. Graph Databases. 2015. O'Reilly Media.
3. Kleppmann M. Designing data-Intensive Applications. 2017. O'Reilly Media.
4. Burns B. Designing Distributed Systems. 2018. O'Reilly Media.
5. Nurkiewicz T, Christensen B. Reactive Programming with RxJava. 2017. O'Reilly Media.

D.G. Arsenjev, D.E. Baskakov, V.P. Shkodyrev (Peter the Great St.Petersburg Polytechnic University, Saint Petersburg)

Contract network protocols of interaction of cognitive agents in distributed intellectual control systems

The concept of consensus dynamics in the management of structurally complex production processes is developing based on contract network protocols of interaction in the environment of cognitive agents. The inefficiency of the classical management theory of complex organizational structures is demonstrated due to their uncertainty, multitasking and multidimensionality. As an advanced management model, an intelligent multi-agent control system is considered.

Секция 4
ПРИКЛАДНЫЕ ЗАДАЧИ ОБРАБОТКИ ИНФОРМАЦИИ В УПРАВЛЕНИИ

ТЕХНОЛОГИЯ ВИБРОАКУСТИЧЕСКОГО МОНИТОРИНГА ГОРОДСКИХ АВТОБУСОВ С ЭЛЕМЕНТАМИ МУЛЬТИАГЕНТНОСТИ

В. А. Акулов[1]
(*ФГБОУ ВО Самарский государственный технический университет, +79276577455, vladislav.a.akulov@gmail.com*)

Аннотация

Разработана и апробирована технология виброакустического мониторинга городских автобусов, ориентированная на применение на протяжении всего жизненного цикла (приобретение – утилизация). Технология обеспечивает получение высокоточной 3D-модели распределения виброакустических параметров, которая представлена спектральными характеристикамиэлементов конструкции, переходными процессами и динамическими изображениями, характеризующими форму колебаний наружных элементов, частоту и среднюю виброскорость. Технология не имеет ближайших аналогов.

Ключевые слова: виброакустический мониторинг, контактные и бесконтактные методы исследований, 3-D модель городского автобуса.

Введение

Как известно, современный город должен обеспечить комфортный уровень жизни людей[2]. Это требование относится ко всем без исключения городским структурам, включая общественный транспорт. В свою очередь, к транспорту предъявляется ряд специфических требований, в числе которых обеспечение так называемого «виброакустического комфорта». Под этим понимаются строгие ограничения на предельные значения низкочастотных вибраций салонов и шумов, причём как в слышимом, так и инфразвуковом диапазонах. Важно отметить, что названные ограничения накладываются на весь срок эксплуатации.

Если виброакустическое состояние новых автобусов контролируется заводом-изготовителем, то при эксплуатации и техническом обслуживании такой контроль не предусмотрен. Как результат, у автобусов, обладающих небольшим начальным запасом виброакустических параметров по отношению к нормам, по истечении некоторого промежутка времени параметры выходят за допустимые пределы. Пассажиры и водители начинают подвергаться интенсивным вибрациям и воздействию шумов. Следовательно, необходим мониторинг (периодический контроль) виброакустического состояния автобусов с последующим принятием решений по устранению выявленных дефектов. Для этого необходима проблемно – ориентированная технология, основанная на глубоких и всесторонних знаниях механизмов возникновения вибраций и шумов, закономерностей их распространения как по силовой установке, так и по салону, а также по взаимодействию активных и пассивных элементов. Заводская информация для решения перечисленных проблем недостаточна.

Технология мониторинга должна учитывать множество факторов, в том числе, специфику эксплуатации городских автобусов и их конструкции. Прежде всего, автобусы городского типа отличаются значительным жизненным циклом. Срок их эксплуатации до капитального ремонта составляет до десяти лет, а пробег – до 1 млн км. Совершенно очевидно, что со временем свойства агрегатов изменяются, а, вместе с ними - параметры виброакустического состояния. Важными факторами, затрудняющими решение проблемы, являются современные подходы, применяемые заводами-изготовителями при проектировании и изготовлении автобусов. В частности, одним из средств, повышающих потребительские качества, является минимизация расхода топлива, которая достигается снижением общей массы автобуса, начиная от двигателя и заканчивая салоном.

[1] Доктор технических наук, профессор

[2] Примером служат материалы VIII Московского урбанистического форума «Мегаполис будущего. Новое пространство для жизни»

Основным методом снижения массы двигателя является переход от V – образной схемы к рядной, причём малоцилиндровой. Вместо восьми цилиндров (см., например, серию ПАЗ) остаётся 6 или 4 цилиндра. Такой переход создаёт условия для интенсификации низкочастотных вибраций в диапазоне 10 – 100 Гц и необходимости принятия мер по их устранению. Уменьшение числа цилиндров, а, следовательно, возрастание мощности каждого из них, сопровождается увеличением интенсивности силового импульса, обусловленного воспламенением топлива. Так, например, если мощность двигателя составляет 240 л.с, то мощность каждого из восьми цилиндров составляет 30 л. с. В случае 4-х цилиндрового исполнения она удваивается и возрастает до60 л.с. Кроме того, если в первом варианте вспышки топлива происходят относительно часто, через каждые 90 градусов поворота коленчатого вала, то во втором – значительно реже: через 180 градусов. Как результат увеличения интенсивности импульсов и длительности интервалов между ними, а также действия ряда сопутствующих факторов, создаются условия, способствующие упомянутой интенсификации низкочастотных вибраций и шумов. Наиболее простым и эффективным способом их подавления является увеличение массы (точнее момента инерции) маховика. Но такая мера вступает в противоречие с концепцией, предусматривающей снижение массы конструкции. Возникает необходимость в поиске компромиссного решения, что не всегда удаётся.

Вторым средством повышения экономичности автобусов является снижение массы салона. Она достигается отказом от рамной схемы, облегчением силовых стоек (в городе отсутствуют условия для опрокидывания), применением разнообразных пластмассовых, тонкостенных металлических панелей и значительным по площади остеклением. Как следствие, салон превращается в составную конструкцию, существенно отличающуюся по жёсткости элементов. Одна часть салона, как правило, хвостовая, оснащается силовой фермой, необходимой для установки силового агрегата. Другая часть, более протяжённая и располагающаяся спереди, выполняется по безрамной схеме. Как результат такой компоновки, а она становится типовой, собственная частота салона снижается и попадает в область рабочих частот силового агрегата. Кроме того, как следствие отличий жёсткости хвостовой и передней части салона и многообразия пассивных элементов конструкции, формируется сложный спектр собственных частот салона, также расположенных в диапазоне рабочих частот силовой установки. Следовательно, создаются благоприятные условия для возникновения резонансов.

Виброакустическое состояние автобусов зависит от согласованности режимов работы двигателя с автоматической коробки передач (АКПП), прежде всего с гидротрансформатором (ГТ), который выполняет функцию сцепления. Если автобус неподвижен (остановки, светофоры, «пробки»), то вращается только насосное колесо. Тогда обтекание неподвижных лопаток турбинного колеса может сопровождаться пульсациями давления и расхода жидкости, которые, в свою очередь, являются источниками низкочастотных вибраций.

К числу широко известных факторов, влияющих на виброакустический комфорт, относится эффективность опор, на которые устанавливается силовой агрегат. Как правило, их основной элемент изготавливается из резины. Следовательно, по истечении времени и силовых воздействий, сопровождающих эксплуатацию, демпфирующие свойства изменяются. Они, как правило, ухудшаются. Кроме того, долговечность опор существенно зависит от вибрационного состояния силового агрегата. Следовательно, необходимо исключить применение недостаточно сбалансированных силовых установок. Однако это правило нередко нарушается.

Как указывалось, жизненный цикл (пробег и многолетняя эксплуатация) городских автобусов отличается продолжительностью. Кроме того, они являются многорежимными объектами. Так как расстояние между остановками невелико (в среднем 300 – 600 м), схема движения приобретает вид: остановка, разгон, кратковременный крейсерский режим, торможение, стоянка на светофорах и в пробках, снова разгон и т. д. При такой циклограмме двигатель работает в широком частотном диапазоне и при различных скоростных и нагрузочных режимах. Так, например, максимальные значения частоты вращения коленвала превышают минимальные (холостой ход) в 2,5 – 3 раза. АКПП также работает в переменном режиме, осуществляя многократное переключение передач как в прямом, так и обратном направлениях (I → VI; VI →I). Следовательно, двигательная установка становится широкополосным генератором вибраций и шумов, на которые откликаются облегчённые салоны. Как результат, обеспечение требования «виброакустического комфорта» на протяжении всего жизненного цикла именно городских автобусов становится проблематичным.

Проблема осложняется непостоянством массы автобуса за счёт многократного изменения количества пассажиров как в течение рейса, так и от рейса к рейсу. Кроме того, при движении поступают возмущения от дорожного покрытия. Таким образом, необходима специализированная проблемно-ориентированная технология мониторинга виброакустического состояния городских автобусов, которая в настоящее время отсутствует. Её разработке и апробации посвящена настоящая статья.

Методологические основы предлагаемой технологии

Основу предлагаемой технологии виброакустического мониторинга городских автобусов составляют следующие разделы: постановка задачи; выбор и построение системы измерения и регистрации виброакустических параметров; выбор методов обработки и представления информации, обеспечивающих построение трёхмерной модели виброакустического состояния автобусов произвольного производителя и выявление скрытых закономерностей динамического взаимодействия элементов конструкции; формулировка направлений дальнейших исследований.

Постановка задачи. *Цель исследований:* разработка и апробация на полноразмерном типовом автобусе технологии 3-х мерного мониторинга виброакустического состояния городских автобусов произвольного изготовителя на основе глубокой детализации контактного взаимодействия активных и пассивных элементов конструкции

Задачи исследований

- Контактная 3-х мерная виброметрия активных и пассивных элементов конструкции.
- Исследование влияния режимов работы двигателя на виброакустические параметры силового агрегата и типовых элементов салона, включая места пассажиров, кондуктора и водителя.
- Оценка влияние загрузки салона на виброакустическое состояние исследуемого автобуса
- Лазернаявиброметрия внешних поверхностей автобусов с определением пространственной формы колебаний (свыше 100 точек).
- Исследование переходных процессов при дозированном импульсном воздействии на переднюю подвеску.
- Исследование переходных процессов контактными методами при резком снижении подачи топлива в диапазоне «номинальный режим – холостой ход».
- Исследование переходных процессов, возникающих в остеклении, при дозированном импульсном воздействии.
- Измерение звукового давления внутри салона в инфразвуковом и звуковом диапазонах
- Обобщение полученных данных и формулировка выявленных скрытых закономерностей
- Формулировка рекомендаций по улучшению виброакустического состояния автобусов в течение всего жизненного цикла.
- Формулировка направлений дальнейших исследований

Объект исследований

Рейсовый автобус марки МАЗ-206 с пробегом 220000 км и временем эксплуатации 10 лет.

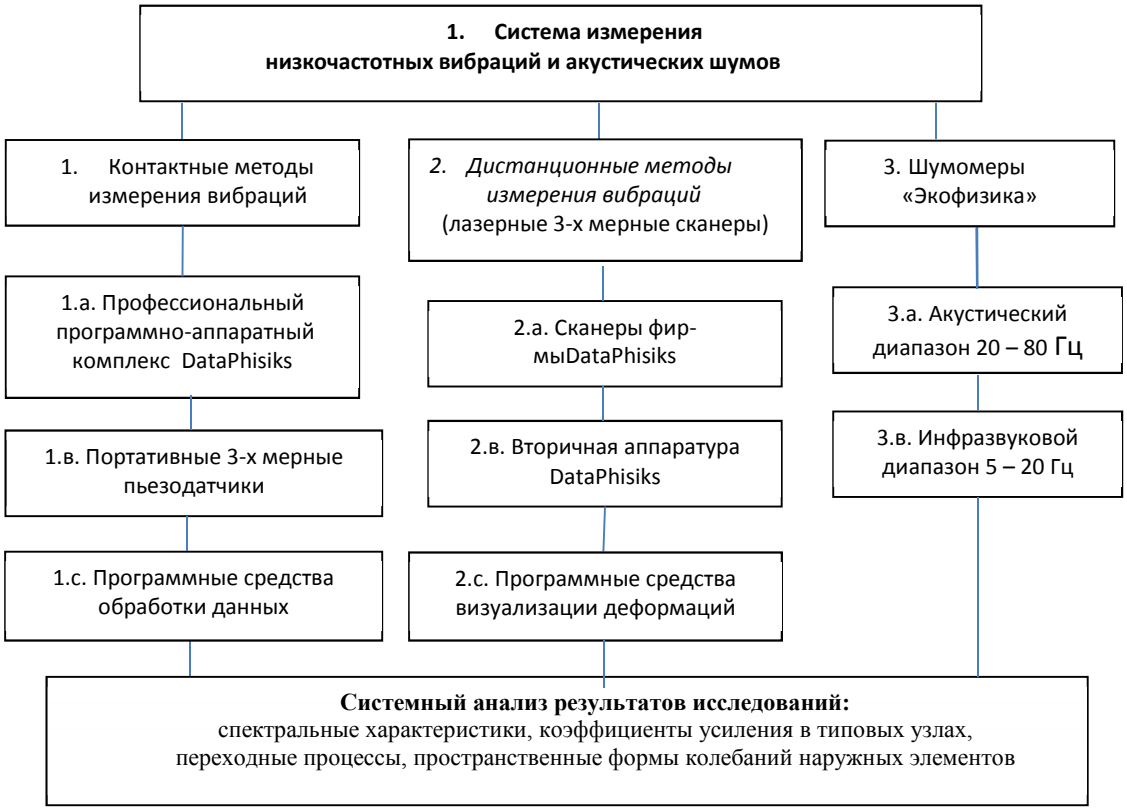
Построение информационно-измерительной системы. Модель состава системы измерений представлена на рис. 1. Как следует из постановки задачи, в технологии предусмотрено применение как контактных, так и бесконтактных методов измерения низкочастотных вибраций, что необходимо для обеспечения высокой достоверности результатов. В связи с этим были применены две системы, отличающиеся принципом действия. Контактная виброметрия осуществлялась комплексом DataPhisiks (рис. 1, группа 1). Учитывая необходимость построения пространственной модели виброакустического состояния были применены 3-х координатные датчики.

С целью выявления маршрутов распространения вибраций и определения направленности взаимодействия элементов силовой установки и салона, начиная от блока двигателя и заканчивая рабочим местом водителя, была введена единая система координат. Ось абсцисс направлялась в продольном направлении салона, ось ординат – вертикально, а ось аппликат в поперечном направле-

нии. Такой подход необходим для установления скрытых механизмов взаимодействия активных и пассивных элементов, выявления проблемных зон и разработки методов гашения вибраций.

Как следует из фотографии, представленной на рис. 1, применялись портативные датчики вибраций, что важно для исключения их влияния на исследуемые элементы конструкции, среди которых были облегчённые панели и стёкла. Бесконтактные методы виброметрии осуществлялись 3-х мерным лазерным сканером (рис. 1, группа 2), в котором реализован метод Допплера для определения скоростных характеристик движущихся поверхностей.

Акустические характеристики салона определялись шумомерами «Экофизика» как в звуковом, так и инфразвуковом диапазонах (рис. 1, группа 3). Следует отметить, что подавление инфразвука является отдельной проблемой, решение которой затруднено рядом факторов. Основной из них - соизмеримость длин волны и продольного размера автобуса, затрудняющая поиск источника шума.





Внешний вид датчика



Общий вид контроллера вибраций Signalstar фирмы DataPhisiks



Общий вид сканирующего лазерного виброметраDataPhisiks
Слева: сканеры,
справа – вторичная аппаратура



Внешний вид шумомера «ЭКОФИЗИКА»

Рис. 1. Обобщённая структурная схема системы измерения

Схема препарирования салона типового автобуса. В соответствии с постановкой задачи, выполнено препарирование автобуса под измерения контактными и бесконтактными датчиками вибраций. На рис. 2 представлен фрагмент схемы препарирования автобуса, отвечающий требованию построения трёхмерной эмпирической модели контактными датчиками. В целях сокращения изложения схему препарирования силового агрегата (точки с № 1– 8) не приводим.

Лазерными методами исследовались вибрации наружных элементов салона – лобового стекла и правой боковой поверхности. На рис. 3 показана фотография лобового стекла автобуса, подготовленного для выполнения сканирования. Для увеличения мощности отражённого сигнала на поверхность стекла были наклеены светоотражающие полосы.

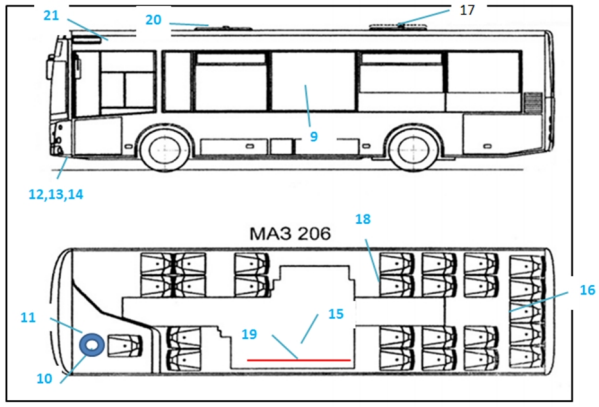


Рис. 2. Фрагмент схемы препарирования автобуса контактными датчиками

Обозначения

9. Боковое стекло салона. 10 Рулевое колесо. 11. Подпятник под ногу водителя. 12. Рулевая колонка. 13. Угловой редуктор. 14. Рулевой механизм. 15. Пол кузова напротив вторых дверей. 16. Задний ряд сидений. 17. Задний люк крыши. 18. Пол кузова место кондуктора. 19. Поручень место для инвалидов. 20. Люк крыши передний. 21. Силовой каркас крыши над водителем



Рис. 3. Фрагмент схемы препарирования салона для лазерного 3-х мерного сканирования. 35 точек доступа

Методы обработки информации. Результаты измерения вибраций контактными методами обрабатывались методами спектрального анализа с построением зависимостей амплитуд вибраций от частоты в диапазоне 0 – 100 Гц. В качестве единицы измерения выбраны перегрузки (G). Результаты представлены в виде графических зависимостей результирующих векторов вибраций и их компонент (распределениям по осям) от частоты. Результаты дистанционного зондирования представлялись видеокадрами, характеризующими форму колебаний, и осреднёнными зависимостями амплитуды виброскорости (мм/с) от частоты.

Основные результаты исследований. Как указывалось, предлагаемая технология обеспечивает решение серии задач, связанных с мониторингом виброакустического состояния городских автобусов. Это определило значительный объём выходной информации. В целях сокращения изложения далее приводятся наиболее общие результаты исследований.

В качестве входного сигнала приняты виброперегрузки блока двигателя. На рис. 4 представлена 3-х мерная модель входного сигнала в виде зависимости амплитуды вибраций от частоты (F,Гц) и скорости вращения коленвала двигателя (N, об/мин). Исследованы три режима работы двигателя: N = 750 об/мин (холостой ход), N = 1200 об/мин и N = 2150 об/мин (номинальный режим). Как следует из рис. 4, в эксплуатационном диапазоне частоты вращения имеет место недостаточная уравновешенность двигателя на крайних режимах (отмечено стрелками). Наиболее интенсивные вибрации, т. е. наиболее интенсивное воздействие на салон автобуса зарегистрированы на режиме холостого хода с частотой 20 Гц. Как показал опыт эксплуатации автобусов, именно на этой частоте салон подвержен интенсивным вибрациям. Причём небольшое увеличение оборотов двигателя при трогании автобуса с места сопровождается суще-

ственным снижением интенсивности колебаний. Для объяснения механизма явления рассмотрим рис. 5, на котором представлена двумерная модель входного сигнала в ограниченном диапазоне частот (0 – 25 Гц).

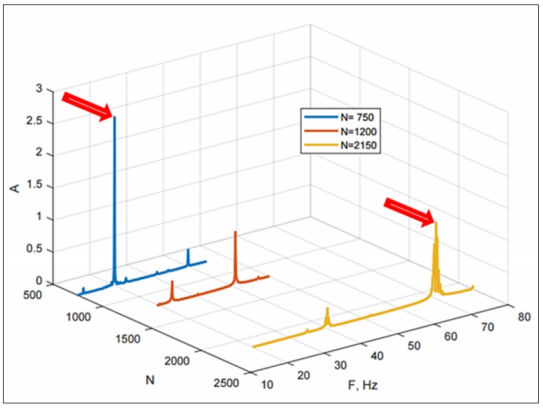


Рис. 4. Трёхмерная модель входного сигнала

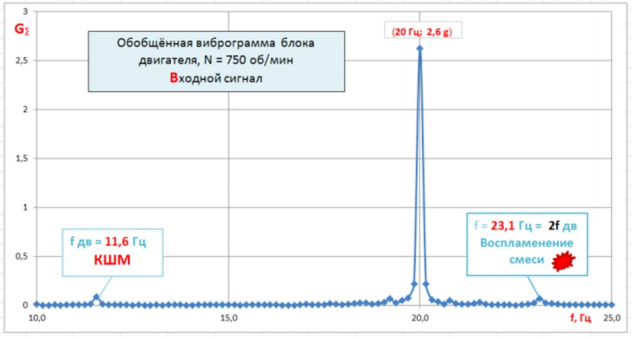


Рис. 5. Зависимость амплитуды виброперегрузок блока двигателя от частоты

Как следует из рис. 5, двигатель создаёт две разновидности возмущений. Первая из них обусловлена работой кривошипно-шатунного механизма (КШМ) и происходит с частотой 11,6 Гц. Вторая разновидность вызвана процессом воспламенения топливной смеси. В соответствии с теорией 4-х цилиндрового двигателя её частота вдвое выше (23,1 Гц).

Анализ показаний других датчиков, расположенных в различных точках салона, свидетельствует о том, что частота первого тона составляет именно 20 Гц. Таким образом, частота вспышек незначительно превышает собственную частоту салона, что и вызывает резонансный эффект. Поскольку частота вспышек превышает собственную частоту, то режиму резонанса соответствует точка на правой, нисходящей ветви резонансной кривой. В соответствии с известными теоретическими положениями с ростом частоты возмущающего воздействия амплитуда вынужденных колебаний снижается. Этим и объясняется вышеупомянутый эффект гашения колебаний при начале движения. Выявленная закономерность позволила предложить в качестве первоочередной меры подавления резонанса некоторое повышение оборотов двигателя (примерно на 80 – 100 об/мин при уровне 750 об/мин).

Как указывалось, одной из задач исследований являлась построение пространственной модели виброакустического состояния автобуса. В качестве фрагмента модели на рис. 6 показано распределение виброперегрузок по частоте на заднем сиденье.

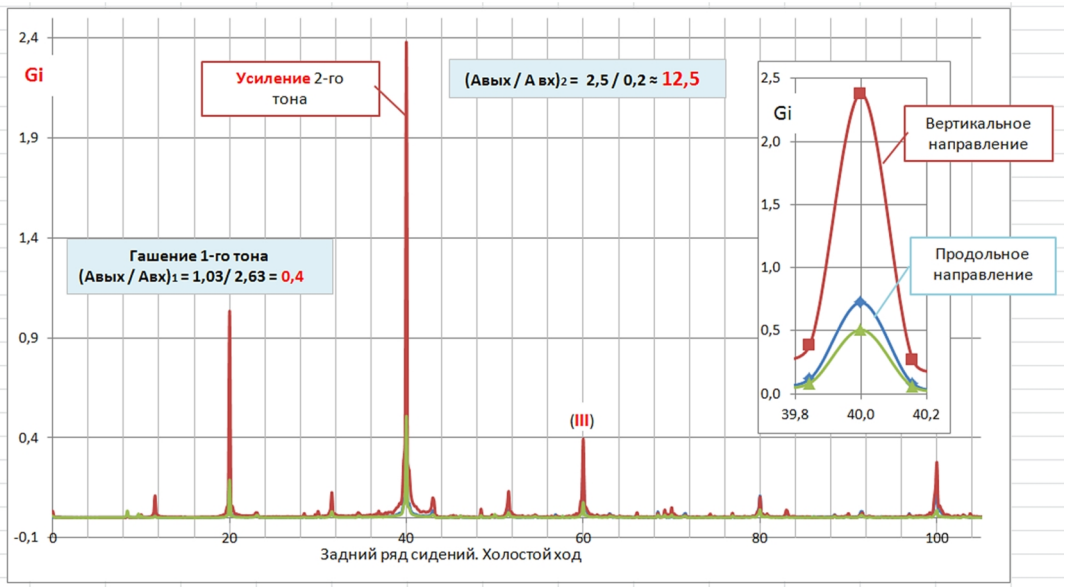
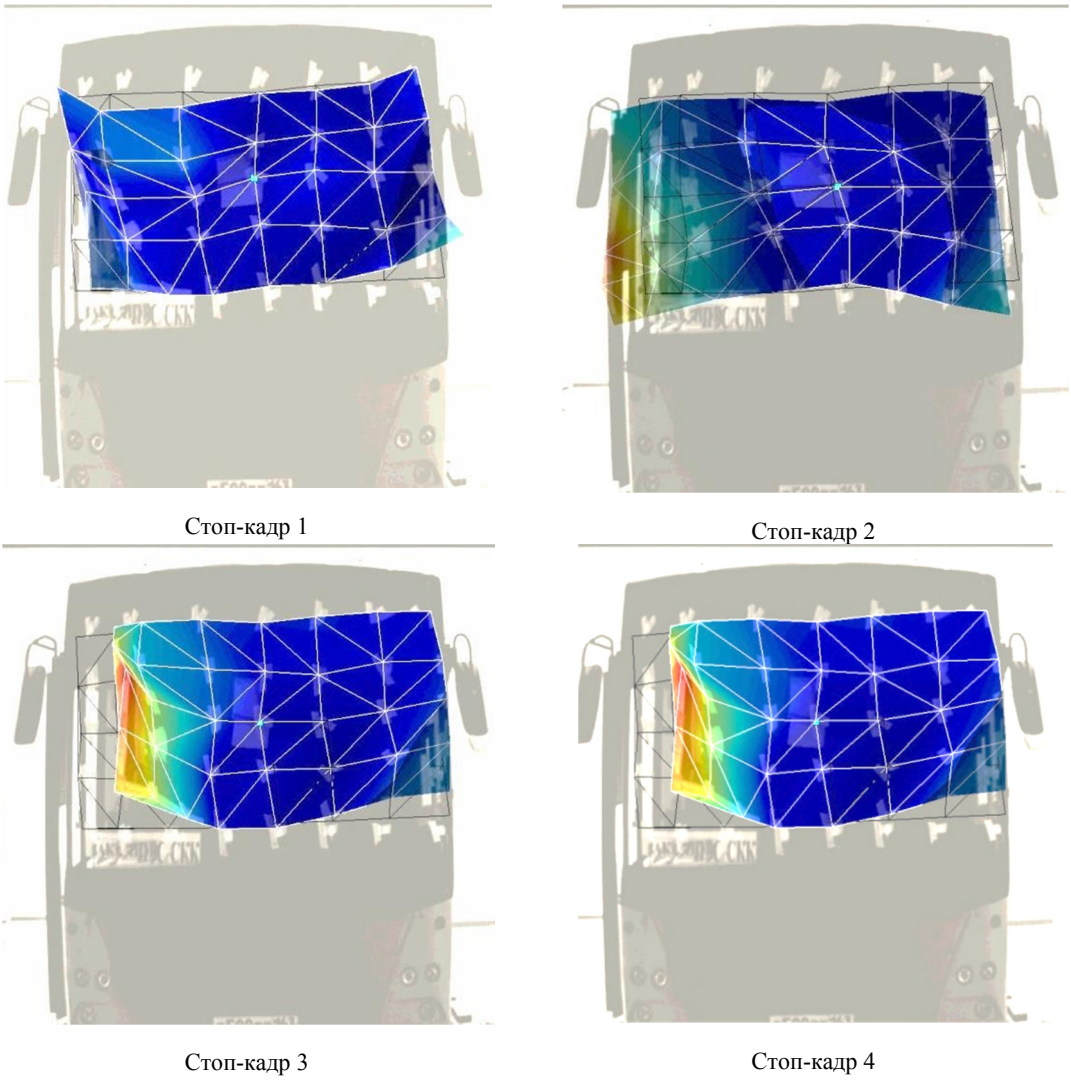


Рис. 6. Спектр виброперегрузок заднего ряда сидений на режиме холостого хода (n = 750 об/мин)

Как следует из рисунка, амплитуда 1-го тона (20 Гц) относительно входного сигнала снизилась от значения 2,63 до 1,03 (коэффициент усиления составил 0,4). В то же время, амплитуда вибраций с частотой 2-го тона (40 Гц) возросла в 12,5 раз, что существенно. В качестве примера на рис. 6 показана структура виброперегрузок (правая область). Как следует из рисунка, на частоте 40Гц основная компонента направлена вертикально (интенсивность около 2,5 единиц), а в продольном направлении виброперегрузки не превышают по величине 0,7 единиц.
Аналогичным образом представляются результаты обработки экспериментов по остальным контактным датчикам.

Как указывалось, спецификой предлагаемой технологии являлся комплексный подход к получению информации, сочетающий контактные и дистанционные методы исследований. Такое сочетание имеет ряд преимуществ. Прежде всего, обеспечивается высокая достоверность результатов. В частности, получено подтверждение скрытых закономерностей динамического взаимодействия элементов конструкции. В качестве примера на рис. 7 приведены результаты лазерного сканирования лобового стекла автобуса (стоп– кадры видиофильма).



Стоп-кадр 1 Стоп-кадр 2

Стоп-кадр 3 Стоп-кадр 4

Рис. 7. Пример визуализации форм колебаний лобового стекла автобусов МАЗ-206 на режиме холостого хода (N = 750 об/мин). С увеличением яркости элемента его скорость возрастает. Шкала скоростей не приводится

Как следует из рисунка, лобовое стекло на режиме холостого хода испытывает весьма существенные деформации. Зарегистрированы скорости вибраций в отдельных точках. Важно отметить, что частота первого тона составила 20Гц, что полностью совпадает с данными, полученными контактными методами.

Заключение

1. Разработана и апробирована в реальных условиях технология виброметрирования пассажирских автобусов, предусматривающая сочетанное применение контактных и бесконтактных методов исследования с целью обеспечения высокой достоверности результатов.
2. В качестве объекта исследований выбран автобус МАЗ-206 с пробегом 220 000 км, что составляет менее трети ресурса. В качестве входного сигнала приняты виброперегрузки блока цилиндров.
3. При всех экспериментах применялась единая система координат: продольное, поперечное и вертикальное направление. Благодаря трёхмерности построена пространственная модель автобуса и выявлены скрытые закономерности динамического взаимодействия его узлов.
4. Установлен механизм повышенного уровня вибраций типового объекта исследований на режиме холостого хода. В результате применения малоцилиндровых двигателей и безрамной схемы автобусов их собственные частоты и частоты возмущений, создаваемых силовой установкой, расположились в одном диапазоне. Как результат совпадения частоты вспышек топливной смеси с собственной частотой первого тона на режиме холостого хода произошёл резонанс.
5. В качестве первоочередной меры устранения вибраций предложено увеличение оборотов двигателя на режиме холостого хода, что согласуется с теоретическими положениями и данными многолетней эксплуатации объектов исследований. Предложенная мера не вызывает затруднений и может осуществляться при очередном техническом обслуживании.
6. Целесообразно продолжение исследований, предназначенных для повышение виброакустического комфорта и надёжности городских автобусов различных модификаций.

V.A. Akulov (FGBOU VO Samara State Technical University, Samara)

Technology of vibroacoustic monitoring of city bus with multi-agency elements

The technology of vibro-acoustic monitoring of city buses, oriented to application throughout the life cycle (acquisition - utilization) has been developed and tested. The technology provides a high-precision 3-D model for the distribution of vibro-acoustic parameters, which is represented by the spectral characteristics of structural elements, transient processes and dynamic images that characterize the shape of the oscillations of the outer elements, frequency and average vibration velocity. Technology has no immediate analogues.

ОЦЕНКА ТОЧНОСТИ НАВИГАЦИИ ВНУТРИ ПОМЕЩЕНИЙ ПО ДАННЫМ BLUETOOTH УСТРОЙСТВ

В. М. Гриняк[1],
(*Владивостокский государственный университет экономики и сервиса, Владивосток, ул. Гоголя, 41, victor.grinyak@gmail.com*)

А. С. Девятисильный[2]
(*Институт автоматики и процессов управления ДВО РАН, Владивосток, ул. Радио, 5, devyatis@iacp.dvo.ru*)

П. А. Цыбанов[3]
(*Дальневосточный федеральный университет, Владивосток, ул. Суханова, 8, scanepson@yandex.ru*)

Аннотация

Работа посвящена проблеме позиционирования мобильных устройств внутри зданий, где ограничена доступность спутниковых навигационных данных. Предлагается модель альтернативной дальномерной маячной навигационной системы на основе Bluetooth датчиков. В основу математической модели задачи навигации положены классические представления системы «состояние-измерение» и метод наименьших квадратов. Такой подход позволяет априорно оценивать характерную точность определения координат объекта, что важно при проектировании навигационной системы рассматриваемого типа. Работа сопровождается результатами натурных экспериментов, подтверждающих расчётные данные.

Ключевые слова: навигация внутри помещений, маячная система, Bluetooth, координаты, скорость.

Введение

В настоящее время навигационные функции мобильных устройств (смартфонов) характеризуются очень широким кругом бытовых потребителей. Кроме определения местоположения они могут решать задачи поиска пути, оценки времени движения между пунктами, предоставления контекстной рекламы и т.п. Местоположение и скорость объекта определяется на основе данных спутниковых навигационных систем второго поколения (GPS, Глонасс), которые достаточно точно делают это в условиях наблюдения достаточного количества спутников, хорошего качества их сигнала и расположения объекта вблизи поверхности Земли [1].

Нахождении объекта внутри помещений нередко ухудшает условия наблюдения: число видимых спутников может быть ограничено, а приём их данных – неустойчивым. В этом случае для определения местоположения объектов внутри помещения могут быть использованы альтернативные источники навигационной информации: инерциальные навигационные средства (акселерометры, гироскопы), точки доступа в интернет Wi-Fi, устройства передачи данных Bluetooth, датчик давления, магнитометр и т.п.

К настоящему времени навигация внутри помещений оформилась в самостоятельный класс задач [2], носящих комплексный характер. Во-первых, они требуют разработки особых математических моделей и алгоритмов фундаментального характера для обеспечения навигационной компоненты. Во-вторых, для их реализации в виде информационных сервисов необходима разработка программных комплексов специальной архитектуры, учитывающей особенности программно-аппаратных платформ мобильных устройств. Среди имеющихся наработок в области навигации внутри помещений можно отметить следующие хорошо зарекомендовавшие себя подходы.

Использование данных инерциальных датчиков. В этом случае с помощью гироскопов и акселерометров мобильного устройства решается либо классическая задача инерциальной нави-

[1] Доктор технических наук, профессор.
[2] Доктор технических наук, заведующий сектором.
[3] Ассистент.

гации, либо её варианты, адаптированные к особенностям имеющихся датчиков и априорных данных о помещении и свойствах движения объекта [3-5]. Достоинством такого подхода является автономность и отсутствие привязки к внешней инфраструктуре. Вместе с тем, устройствам бытового назначения инерциальные измерения доступны лишь с большой погрешностью, что не даёт возможности решать задачу с приемлемой точностью. Это делает указанный подход практически полезным только для специальных устройств, оснащённых навигационными датчиками высокой точности.

Использование данных Wi-Fi/Bluetooth. Первый способ использования таких данных – решение задачи типа трилатерации. В этом случае исходными данными для оценки координат наблюдаемого объекта служат измерения дальностей «объект-датчик», при этом дальность оценивается по уровню принимаемого мобильным устройством сигнала [6, 7]; такой способ аналогичен используемым, например, в гидроакустике [8-10]. Второй способ – определение местоположения объекта путём сравнения измеряемых уровней сигнала от окружающих датчиков с заранее измеренными значениями, привязанными к карте помещения (карта уровня сигналов, радиоотпечаток) [11-14]. К достоинствам подхода можно отнести возможность использования уже развёрнутых сетей передачи данных (Wi-Fi). Недостатком является высокая погрешность измерения уровня сигнала, дискретность данных карт уровня сигналов, принципиально ограничивающая точность навигации и необходимость предварительной калибровки и настройки (обучения) системы.

Имеются сведения о попытках разработки других подходов, например, системы оптического типа, когда сравниваются данные сканирования помещения и изображения с камеры смартфона [15] или системы, основанной на навигации по магнитному полю с помощью компаса смартфона [16]. Известны попытки комбинирования в одной системе позиционирования нескольких подходов одновременно [17-19].

Использование для задач навигации внутри помещений Bluetooth устройств имеет целый ряд преимуществ. Прежде всего, это построение инфраструктуры на базе недорогого и распространённого оборудования, не требующего специальных профессиональных навыков персонала и возможность произвольного конфигурирования датчиков на стенах помещения в силу их малого размера и автономности. Кроме того, дальность видимости Bluetooth устройств невелика, а их сигнал практически не проходит сквозь капитальные стены, что полностью решает «проблему этажа», актуальную, например, для спутниковых навигационных систем [1].

Предлагаемый доклад посвящен результатам исследования модели задачи навигации внутри помещений по данным Bluetooth устройств, основанной на измерениях уровня сигнала, интерпретирующего дальность «объект-датчик». Целью исследования является оценка характерной точности определения координат объекта в различных типовых навигационных ситуациях и перспектив реализации подхода для тех или иных практических приложений.

Модели и методы

Передаваемые Bluetooth устройством данные содержат довольно ёмкую информацию о его работе [20]. С точки зрения обсуждаемой задачи наиболее важными элементами данных являются уникальный идентификатор наблюдаемого устройства и относительный уровень мощности принимаемого от него сигнала (единицей измерения, используемой в смартфонах, обычно служит дБм). Уровень принимаемого сигнала связан с дальностью «объект-датчик» следующим известным соотношением [21]:

$$u = u_* - 10 \cdot n \cdot \log_{10}(d / d_*).$$

Здесь u – уровень сигнала, принимаемого мобильным устройством (объектом) на расстоянии d от Bluetooth устройства (датчика), u_* – уровень сигнала, измеренный на расстоянии d_* от устройства, n – коэффициент, учитывающий распространение сигнала в окружающей среде. Значения параметров d_* и u_* – калибровочные расстояние и уровень сигнала – являются техническими характеристиками конкретного устройства, значение коэффициента затухания n также считается известным. В принципе, эти значения могут быть определены при калибровке устройства.

Введём правую ортогональную систему координат *xy*, связанную с помещением. Пусть имеется *N* датчиков, «видимых» мобильным устройством. Дальность «объект-датчик» связана с координатами объекта и датчика следующим образом:

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}\ ,\ i = \overline{1, N},$$

где *x*, *y* – координаты объекта, x_i , y_i – координаты *i*-го датчика.

Уравнения измерений будут иметь вид:

$$u_i(t_k) = u_{*i} - 10 \cdot n \cdot \log_{10}\left(\sqrt{(x(t_k) - x_i)^2 + (y(t_k) - y_i)^2}\ / d_{*i}\right) + \eta_i(t_k)\,, \qquad (1)$$

где $u_i(t_k)$ – уровень сигнала *i*-го датчика, принимаемого на объекте в момент времени t_k , x_i , y_i – координаты *i*-го датчика, $x(t_k)$, $y(t_k)$ – координаты объекта в момент времени t_k , u_{*i} и d_{*i} – калибровочные значения уровня сигнала и расстояния *i*-го датчика, $\eta_i(t_k)$ – инструментальная погрешность измерения *i*-го датчика, отнесённая к моменту времени t_k , $k = \overline{1, M}$, *M* – число измерений от каждого датчика.

Положим, что движение объекта описывается следующей кинематической моделью:

$$\begin{aligned} x(t_{k+1}) &= x(t_k) + v_x(t_k)(t_{k+1} - t_k), \\ y(t_{k+1}) &= y(t_k) + v_y(t_k)(t_{k+1} - t_k). \end{aligned} \qquad (2)$$

Здесь $x(t_k)$, $y(t_k)$ – координаты объекта в момент времени t_k , $v_x(t_k)$, $v_y(t_k)$ – компоненты вектора скорости объекта.

Линеаризация уравнения (1) приводит его к виду:

$$\begin{aligned} \delta u_i(t_k) = {} & \frac{-10 \cdot n \cdot d_{*i} \cdot (x(t_k) - x_i)}{\ln(10) \cdot \left((x(t_k) - x_i)^2 + (y(t_k) - y_i)^2\right)} \delta x(t_k) + \\ & + \frac{-10 \cdot n \cdot d_{*i} \cdot (y(t_k) - y_i)}{\ln(10) \cdot \left((x(t_k) - x_i)^2 + (y(t_k) - y_i)^2\right)} \delta y(t_k) + \eta_i(t_k), \end{aligned} \qquad (3)$$

где $\delta x(t_k)$, $\delta y(t_k)$ – погрешности априорных представлений о координатах объекта, $\delta u_i(t_k)$ – невязка измерения.

Популярными методами решения задач типа (2, 3) являются динамические алгоритмы статистически оптимального оценивания, усваивающие измерения последовательно, по мере поступления данных (алгоритм Калмана, модальные алгоритмы [22, 23]). Особенности программной реализации рассматриваемой задачи с накоплением и совместной обработкой данных измерений за некоторый период времени делают возможным альтернативный подход – сведение её к методу наименьших квадратов, в пользу которого и сделан выбор в настоящей работе.

Запишем уравнения (3) с учётом модели эволюции состояния системы (1):

$$\begin{aligned} \delta u_i(t_k) = {} & \frac{-10 \cdot n \cdot d_{*i} \cdot (x(t_k) - x_i)}{\ln(10) \cdot \left((x(t_k) - x_i)^2 + (y(t_k) - y_i)^2\right)} \delta x(t_*) + \\ & \frac{-10 \cdot n \cdot d_{*i} \cdot (x(t_k) - x_i)}{\ln(10) \cdot \left((x(t_k) - x_i)^2 + (y(t_k) - y_i)^2\right)} (t_k - t_*) \delta v_x(t_*) + \\ & + \frac{-10 \cdot n \cdot d_{*i} \cdot (y(t_k) - y_i)}{\ln(10) \cdot \left((x(t_k) - x_i)^2 + (y(t_k) - y_i)^2\right)} \delta y(t_*) + \\ & + \frac{-10 \cdot n \cdot d_{*i} \cdot (y(t_k) - y_i)}{\ln(10) \cdot \left((x(t_k) - x_i)^2 + (y(t_k) - y_i)^2\right)} (t_k - t_*) \delta v_y(t_*) + \eta_i(t_k), \end{aligned}$$

или, в общем случае:

$$\delta Z = H \delta s(t_*) + \xi\,. \qquad (4)$$

Здесь δZ – полный вектор (длины $M \times N$) невязок измерений на интервале наблюдения, $\delta s(t_*)$ – вектор погрешностей априорных представлений о состоянии системы (координатах и скоростях

объекта) в выбранный момент времени t_*, H – матрица коэффициентов, ξ – вектор приведённых погрешностей измерений. Целью решения задачи (4) является оценка вектора $s(t_*)$ по имеющимся измерениям уровней сигнала $u_i(t_k)$. Эта оценка может быть выполнена, например, путём сведения задачи к системе линейных алгебраических уравнений:

$$H^T Q^{-1} \delta Z = H^T Q^{-1} H \delta s(t_*) \qquad (5)$$

и её решения известными способами. Здесь $Q = \mathrm{mean}[\xi\xi^{\mathrm{T}}]$ – ковариационная матрица случайного вектора ξ, она определяется свойствами инструментальных погрешностей измерений $\eta_i(t_k)$ и считается известной. Матрица $D = (H^{\mathrm{T}} Q^{-1} H)^{-1}$ определяет при этом дисперсию погрешности оценки вектора состояния $\delta\hat{s}(t_*)$, так что $D = \mathrm{mean}[(\delta\hat{s}(t_*) - \delta s(t_*))(\delta\hat{s}(t_*) - \delta s(t_*))^{\mathrm{T}}]$. mean – оператор математического ожидания.

В рассматриваемой задаче матрица D довольно точно описывает вероятностные свойства как её линеаризованного варианта (2, 3), так и основного, исходно нелинейного (2, 1). Как показывают приведённые ниже результаты исследования, значения соответствующих коэффициентов матрицы D вполне позволяют оценить характерную точность определения координат объекта при различных конфигурациях системы датчиков, используемых для навигации внутри помещения.

Результаты численного моделирования

Передаваемые Bluetooth устройством данные содержат довольно ёмкую инф Численное исследование задачи проводилось для следующих значений параметров: инструментальная погрешность измерений датчиков имеет нормальное распределение с дисперсией 25дБм^2; частота обновления данных об уровне сигнала равна 0.07с, что характерно для современных смартфонов; количество измерений M от каждого датчика равно 28, что соответствует длине интервала наблюдения приблизительно 2 секунды.

На рис. 1 и 2 показано моделируемое расположение датчиков; имитируются следующие типичные ситуации: коридор шириной 3м (рис. 1) и холл или комната размером 10×10м (рис. 2).

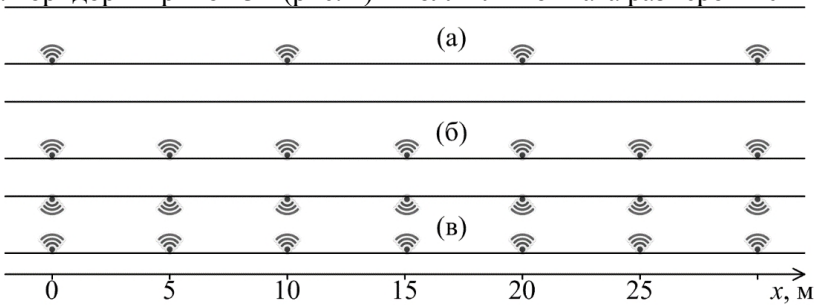


Рис. 1. Варианты расположения датчиков в помещении типа «коридор»

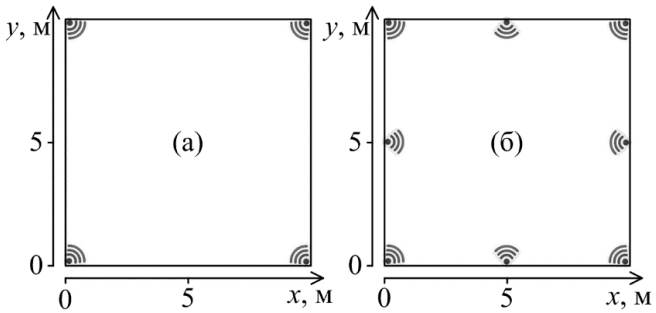


Рис. 2. Варианты расположения датчиков в помещении типа «холл» или «комната»

Датчики «наклеены» на стены в различных конфигурациях: с одной стороны коридора (рис. 1а, 1б), с двух сторон коридора (рис. 1в), в углах комнаты (рис. 2а), в углах и на стенах комнаты

(рис. 2б). При этом считается, что датчики расположены в одной плоскости с движущимся объектом (или достаточно близко к этой плоскости), что даёт возможность ограничиться только двумерным вариантом модели рассматриваемой задачи.

Будем рассматривать два варианта вектора состояния $s(t_k)$: с компонентами, включающими координаты и скорости объекта (кинематическая задача) и с компонентами, включающими только координаты (статическая задача).

На рис. 3а показаны вычисленные согласно (5) значения средней погрешности оценивания координат объекта σ_{xy} в кинематической задаче, так что

$$\sigma_{xy} = \sqrt{D_{xx} + D_{yy}} ,$$

где D_{xx} и D_{yy} – соответствующие диагональные коэффициенты матрицы D; моделируется, что объект находится (движется) по середине коридора (рис. 1). Сплошная линия соответствует конфигурации датчиков рис. 1в, пунктир – рис. 1б, точки – 1а.

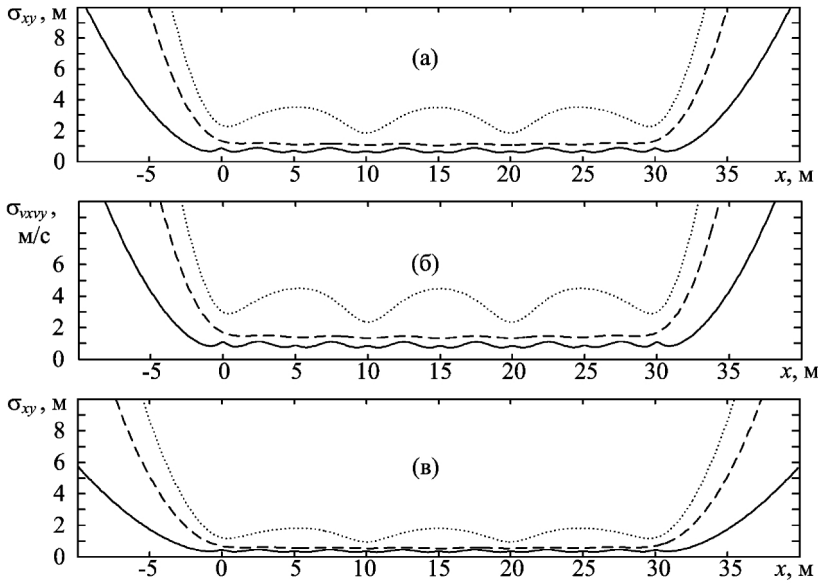


Рис. 3. Погрешность определения координат и скоростей объекта в помещении типа «коридор»

Видно, что конфигурация системы датчиков и их расположение относительно объекта существенно влияет на величину погрешности оценивания. Так, при нахождении объекта дальше 7 метров от системы датчиков (области графика $x < -7$ и $x > 37$) величина σ_{xy} становится неприемлемой во всех случаях (более 5 метров). Если объект находится «внутри» системы датчиков (область графика $0 < x < 30$), величина σ_{xy} равна от 1м (в случае, если датчики расположены на расстоянии 5м друг от друга) до 3.5 метров (если датчики расположены на расстоянии 10м друг от друга). Расположение датчиков с двух сторон также улучшает условия наблюдения.

На рис. 3б показаны значения средней погрешности оценивания скоростей объекта $\sigma_{v_x v_y}$ в кинематической задаче, так что

$$\sigma_{v_x v_y} = \sqrt{D_{v_x v_x} + D_{v_y v_y}} ,$$

где $D_{v_x v_x}$ и $D_{v_y v_y}$ – соответствующие диагональные коэффициенты матрицы D; объект также движется по середине коридора. Величина $\sigma_{v_x v_y}$ принимает значения от 1 до 4 м/с.

На рис. 3в показаны значения средней погрешности оценивания координат объекта σ_{xy} в статической задаче при нахождении объекта в коридоре. Видно, что при нахождении объекта

«внутри» системы датчиков величина σ_{xy} равна от 0.5 до 2 метров, что, как и ожидалось, существенно лучше, чем в кинематической задаче (рис. 3а).

На рис. 4 показаны вычисленные согласно (5) значения средней погрешности оценивания координат объекта σ_{xy} в кинематической задаче при нахождении (движении) объекта в комнате (рис. 2). Рис. 4а соответствует конфигурации датчиков рис. 2а, а рис. 4б – конфигурации 2б. Величина σ_{xy} принимает значения от 1.9 до 2.5 метров для случая четырёх датчиков и от 0.9 до 1.2 метра для комнаты с восемью датчиками.

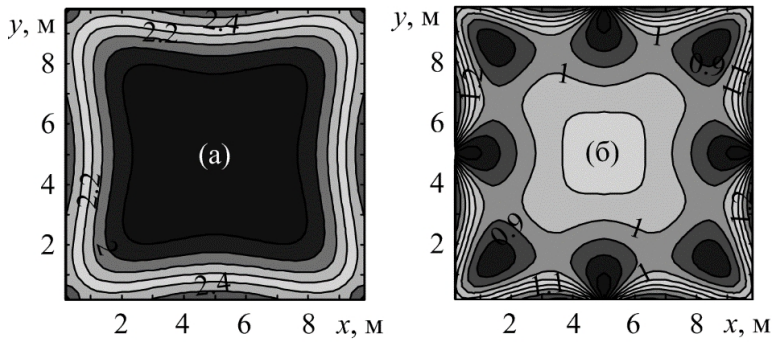


Рис. 4. Погрешность определения координат в помещении типа «комната» при оценке координат и скоростей объекта

Аналогичные данные приведены на рис. 5 – показаны средние погрешности оценивания скоростей объекта $\sigma_{v_x v_y}$ в кинематической задаче и на рис. 6 – показаны значения величины σ_{xy} в статической задаче. В последнем случае они примерно в два раза меньше, чем в кинематической задаче.

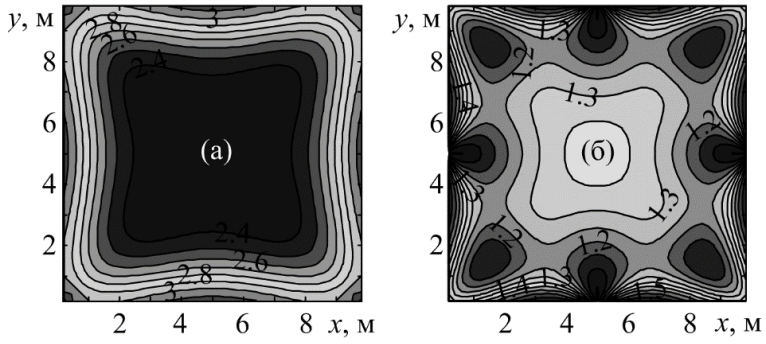


Рис. 5. Погрешность определения скорости в помещении типа «комната» при оценке координат и скоростей объекта

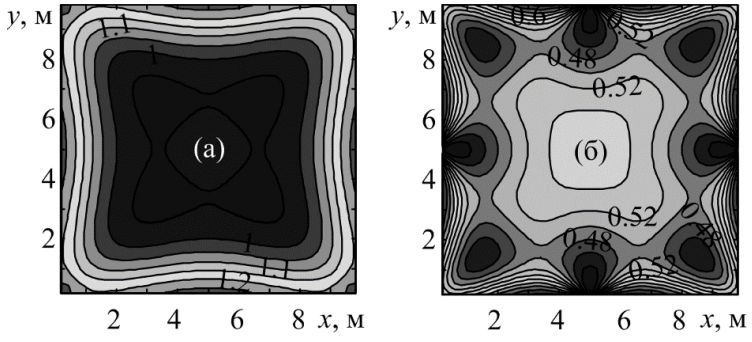


Рис. 6. Погрешность определения координат в помещении типа «комната» при оценке только координат объекта

В целом ожидаемую погрешность оценивания координат следует признать приемлемой для большого числа практических приложений, погрешность же оценивания скорости (рис. 3б, 5) довольно велика.

Результаты натурного эксперимента

При проведении натурных исследований использовались Bluetooth маяки SKYLAB Beacon VG01 [24] и мобильный телефон Samsung Galaxy S8 на платформе Android. На рис. 7 показан результат оценки координат объекта в помещении типа «комната» при решении кинематической задачи.

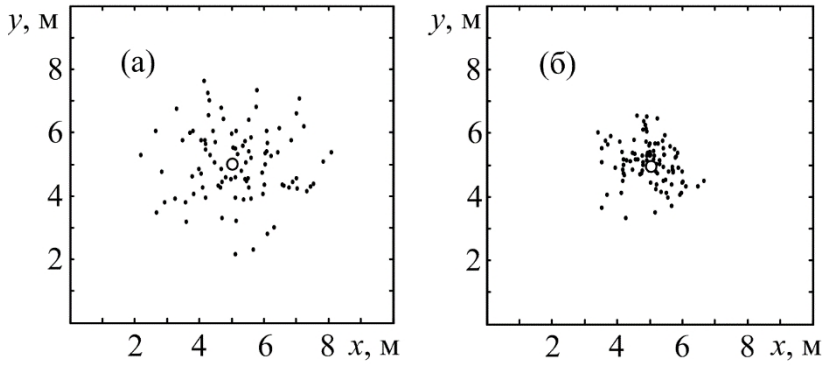


Рис. 7. Результаты оценки координат объекта в помещении типа «комната» при оценке координат и скоростей объекта

Точками показаны оценки местоположения объекта (100 точек – 100 последовательных решений задачи), при этом объект находился в середине комнаты (выколотый кружок). Рис. 7а соответствует случаю четырёх маяков по углам комнаты (рис. 2а), рис. 7б – случаю восьми маяков (рис. 2б).

На рис. 8 показан аналогичный результат оценки координат объекта при решении статической задачи.

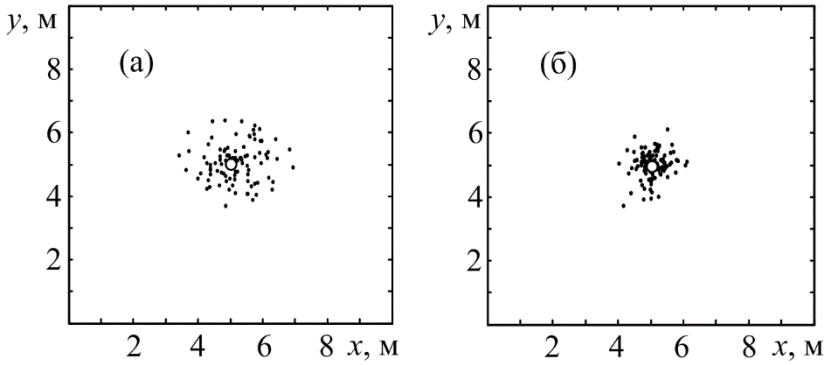


Рис. 8. Результаты оценки координат объекта в помещении типа «комната» при оценке только координат объекта

Из результатов проведённых испытаний можно сделать вывод о том, что погрешность оценки координат объекта в натурном эксперименте полностью соответствует предсказанной расчётами. Для этого достаточно соотнести данные рисунков 5 и 7, 6 и 8; видно, что погрешность оценки координат укладывается в круг с радиусом $2\sigma_{xy}$. Таким образом, теоретические данные о средней погрешности оценивания координат и скоростей объекта, полученные для линеаризованной задачи (из уравнения (5)), достаточно близки к реальным свойствам исходной задачи (1, 2), что подтверждает возможность их использования для анализа разрешимости при проектировании навигационной системы рассматриваемого типа.

Обсуждение результатов

Данные расчётов и натурных экспериментов показывают, что характер навигационных измерений, доступных для устройств типа Bluetooth, даёт возможность качественного решения задачи наблюдения на характерных дальностях «объект-датчик» не превышающих 5-7 метров. При этом важны также условия наблюдения – взаимное расположение датчиков и объекта; наилучший результат обеспечивается в том случае, если датчики располагаются как бы «окружая» объект.

Точность решения статической задачи существенно выше (примерно в 2 раза), чем кинематической. С другой стороны, игнорирование в модели задачи скорости объекта при его реальном движении приводит к её неудовлетворительному решению. Поэтому при разработке приложения, осуществляющего навигацию мобильного устройства, следует обязательно реализовать идентификацию модели движения. Для покоящихся на интервале наблюдения объектов ограничиваться статической моделью задачи, для движущихся – использовать кинематическую модель. Создавая приложение необходимо также учитывать ресурсоёмкость операции чтения данных навигационных датчиков, при этом желательна насколько возможно высокая частота обновления данных их внутреннего буфера.

Общей проблемой всех задач, решаемых методом линеаризации около опорного решения, является локальность оценок их свойств. Проведённые натурные эксперименты показывают, что в рассматриваемой задаче свойства линеаризованной модели достаточно хорошо описывают исходную (нелинейную) задачу.

Построение навигационной системы рассматриваемого типа всегда связано с калибровкой датчиков и их привязкой к местной локальной системе координат (выставкой системы). Такая привязка может быть осуществлена различными известными способами, в том числе описанными в работах авторов, опубликованных ранее [25, 26].

Заключение

Результаты настоящего исследования позволяют оптимистично оценить возможность построения навигационной системы на основе Bluetooth маяков с использованием классических модельных представлений маячных дальномерных систем. При характерном расстоянии между маяками и объектом 5м реально достижимая точность определения координат объекта составляет 1-2м, что достаточно для традиционных задач навигации внутри зданий (построение пути, контекстная реклама). Дальнейшее повышение точности навигации может быть связано с уменьшением случайной компоненты инструментальных ошибок измерений путём идентификации модели сигнала и применения различных алгоритмов фильтрации, чему авторы планируют посвятить отдельное исследование. Представляется, что улучшение точности навигации до значений порядка 0.5м позволит качественно расширить возможный круг решаемых задач (навигация автономных робототехнических устройств, идентификация аномалий движения в потоке людей и т.п.).

ЛИТЕРАТУРА

1. Шебшаевич В.С. Сетевые спутниковые радионавигационные системы. – М.: Радио и связь, 1993. 408с.
2. International Conference on Indoor Positioning and Indoor Navigation [Электронный ресурс] – Режим доступа http://ipin-conference.org/ (01.04.2018).
3. Дэвидсон П., Киркко-Яаккола М., Коллин Ю.З, Такала Я. Навигационный алгоритм с использованием планов зданий и данных автономных датчиков // Гироскопия и навигация. – 2015. – №1. – с. 29-42.
4. Кроненветт Н., Руппельт Я., Троммер Г.Ф. Прецизионное позиционирование пешехода в помещении на основе контроля за стадиями его походки // Гироскопия и навигация. – 2017. – №1. – с. 33-48.
5. Емельянцев Г.И., Степанов А.П., Блажнов Б.А. О решении навигационной задачи для летательных аппаратов с использованием инерциального модуля на микромеханических датчиках и наземных радиоориентиров // Гироскопия и навигация. – 2017. – №1. – с. 3-17.
6. Щёкотов М.С., Кашевник А.М. Сравнительный анализ систем позиционирования смартфонов в помещениях // Труды СПИИРАН. – 2012. – №4. – с. 459-471.
7. Namiot D. On indoor positioning // International Journal of Open Information Technologies. – 2015. – №3. – p. 23-26.
8. Матвиенко Ю.В., Ваулин Ю.В., Каморный А.В. Градуировка навигационных средств подводных роботов // Подводные исследования и робототехника. – 2015. – №1. – с. 16-22.

9. Васильев К.К., Бобков А.В. Динамическое использование ориентиров для оценки координат автономных необитаемых подводных аппаратов // Информационно-измерительные и управляющие системы. – 2017. – №12. – с. 11-14.
10. Мартынюк М.В., Генералов А.В., Наумов С.С., Залетнов С.Е., Дмитриев Д.В., Тимофеева О.П. Разработка системы локального акустического позиционирования на базе персонального компьютера // Измерительная техника. – 2015. – №10. – с. 68-71.
11. Воронов Р.В., Мощевикин А.П. Применение условной энтропии при формировании рекомендаций по размещению базовых станций в локальных системах позиционирования // Информационные технологии. – 2014. – №10. – с. 11-16.
12. Мурашко И.А., Храбров Д.Е. Методика локального позиционирования на основании Wi-Fi-сети университета // Веснік Гродзенскага дзяржаўнага ўніверсітэта імя Янкі Купалы. Серыя 2: Матэматыка. Фізіка. Інфарматыка, Вылічальная тэхніка і кіраванне. – 2015. – №2. – с. 119-127.
13. Воронов Р.В., Малодушев С.В. Динамическое создание карт уровня Wifi-сигналов для систем локального позиционирования // Системы и средства информатики. – 2014. – №1. – с. 80-92.
14. Ассур О.С., Филаретов Г.Ф. Разработка комплексного метода позиционирования объектов по данным беспроводных сетей Wi-Fi и устройств BLE (Bluetooth Low Energy) // Известия Института инженерной физики. – 2015. – №2. – с. 2-10.
15. Yang S.H., Jeong E.M., Kim D.R., Kim H.S., Son Y.H., Han S.K. (2013). Indoor three-dimensional location estimation based on LED visible light communication // Electronics Letters. – 2013. – №49(1). – p. 54-56.
16. Magnetic field anomalies [Электронный ресурс] – Режим доступа http://www.gizmag.com/magneticanomaly-indoor-positioning/23253/ (01.04.2018).
17. Петрова О.А., Табунщик Г.В., Ван Мероде Д. Метод определения текущего расположения в системах позиционирования и навигации внутри помещения // Электротехнические и компьютерные системы. – 2017. – №25. – с. 270-278.
18. Овсянников А.А., Новиков П.А. Модели реализации навигации внутри помещения при помощи анализа беспроводных источников данных // Компьютерные инструменты в образовании. – 2015. – №4. – с. 37-51.
19. Retscher G. Test and Integration of Location Sensors for a Multi-Sensor Personal Navigator // Journal of Navigation. – 2007. – Vol. 60. – №1. – p. 107-117.
20. Протоколы Bluetooth [Электронный ресурс] – Режим доступа http://www.gaw.ru/html.cgi/txt/doc/Wireless/bluetooth/7.htm (01.04.2018)
21. 21. Motley A.J., Keenan J.M.P. Personal communication radio coverage in buildings at 900 MHz and 1700 MHz // Electronics Letters. – 1988. – Vol. 24. – №12. – p. 763-764.
22. Степанов О.А. Основы теории оценивания с приложениями к задачам обработки навигационной информации. Часть 1. Введение в теорию оценивания. – СПб.: ГНЦ РФ ОАО «Концерн «ЦНИИ «Электроприбор», 2010. – 509с.
23. Степанов О.А. Основы теории оценивания с приложениями к задачам обработки навигационной информации. Часть 2. Введение в теорию фильтрации. – СПб.: ГНЦ РФ ОАО «Концерн «ЦНИИ «Электроприбор», 2012. – 417с.
24. Skylab Beacon virgo|beacon for indoor position [Электронный ресурс] – Режим доступа http://www.skylabmodule.com/beacon-virgo/ (01.04.2018)
25. Grinyak V.M., Devyatisil'nyj A.S. Dynamic adjustment of multiposition observing system with respect to trajectory measurements // Journal of Computer and Systems Sciences International. – 1999. – Vol. 38. – №1. – p. 124-130.
26. Гриняк В.М., Гриняк Т.М., Иваненко Ю.С. Динамическая выставка многопозиционной системы наблюдения // Моделирование, оптимизация и информационные технологии. – 2017. - №3. – с. 12.

V.M. Grinyak (Vladivostok State University of Economics and Service, Vladivostok), **A.S. Devyatisilnyi** (Institute of Automation and Control Processes, Far Eastern Branch of Russian Academy of Sciences, Vladivostok), **P.A. Tsibanov** (Far Eastern Federal University, Vladivostok)

Feasibility Study of Indoor Navigation System with Bluetooth Beacons

This work is devoted to research of possibility to use Bluetooth devices for indoors navigation. There is a common model definition for the object movement at the superposition field of set of “visible” Bluetooth beacon signals. For solutions authors relay on linear approximation with applying least squares method. The research includes numerical experiment results for defined set of physical parameters of the system such as an error of evaluation of signal level, signal sampling frequency typical for modern smartphones. There are results of 2D navigation modeling for 2 types of indoor geometry (big square room and long narrow corridor) with different configurations of navigation system (location of Bluetooth transmitters on the walls). The research covers 2 main problems: static and kinematic. The paper also presents results of real experiments for static and kinematic problems with using SKYLAB Bacon VG01 Bluetooth transmitters and smartphone Samsung Galaxy S8. Real experiments confirmed that math model with linear approximation defined by authors is usable for solving indoors navigation problems using Bluetooth signal.

МОДЕРНИЗАЦИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ОБЛАСТИ ОБРАБОТКИ ИЗОБРАЖЕНИЙ

М. В. Харинов
(*СПИИРАН, 199178, С.-Петербург, 14 линия, 39, khar@iias.spb.su*)

Аннотация

В докладе формализуется задача локализации и детектирования объектов на цветовом изображении. Обсуждаются стереотипы, сложившиеся в области обработки цифровых изображений, которые препятствуют активному применению классического кластерного анализа. Для характеристики наличия объектов на изображении вводится «мера яркостной неоднородности участков изображения», которая оценивается как абсолютная величина приращения суммарной квадратичной ошибки при разделении надвое кластеров пикселей. Предлагается способ детектирования иерархии объединений объектов.

Ключевые слова: цифровое изображение, сегментация, кластеризация пикселей, суммарная квадратичная ошибка

Введение

За более чем полувековую историю своего развития практика постановки задачи детектирования и распознавания объектов на изображении прошла путь от попыток создания больших баз эталонов и обучаемых на примерах программных систем до разработки инструментариев обработки изображений, помогающих инженерам-программистам эффективно разрабатывать приложения для решения текущих прикладных задач (распознавания текстов, лиц, номерных знаков и др.). Не исключено, что очередной вариант требуемого программного инструментария создан в результате выполнения в 2013-2017 годах проекта PPAML (Probabilistic Programming for Advancing Machine Learning) американского агентства DARPA (Defense Advanced Research Projects Agency) и, со временем, станет общедоступным, так как снижает стоимость разработки прикладного программного обеспечения (ПО). Разумной альтернативой вероятностному подходу PPAML к разработке ПО для создания обучаемых программ является детерминированный подход к созданию программ, которые настраиваются на детектирование «объектов интереса» с помощью предусмотренной системы управляющих параметров. Обсуждаемая в настоящей статье модель детектирования объектов развивается в рамках детерминированного подхода к распознаванию цифрового изображения.

На входе модели рассматривается матрица трехзначных целочисленных значений яркости цветовых пикселей, которые преобразуются в *вещественные* значения посредством усреднения внутри обрабатываемых кластеров пикселей или сегментов изображения. На выходе модель поддерживает разбиения изображения на непересекающиеся кластеры пикселей, в частности, сегменты изображения, которые рассматриваются в качестве *объектов*, детектируемых компьютером для дальнейшего анализа признаков и распознавания. Вычисления сводятся к первоначальному построению и запоминанию иерархической последовательности (иерархии) кластеров пикселей, из числа которых отфильтровываются объекты при автоматической обработке, или в процессе настройки программной системы.

Целевая иерархия кластеров пикселей задается в модели посредством оптимальных и квазиоптимальных (близких к оптимальным) *кусочно-постоянных приближений* изображения, которые для данного разбиения получаются заполнением кластеров усредненными значениями пикселей и минимально отличаются от изображения по среднеквадратичному отклонению σ (СКО) или *ошибке аппроксимации*, т.е. суммарной квадратичной ошибке $E = 3N\sigma^2$, где N — число пикселей в цветовом RGB– изображении.

Как ни парадоксально, но системное применение кластерного анализа [1, 2] и классических методов минимизации СКО для иерархической кластеризации пикселей пока остается недостаточно исследованным ни в части толкования самих методов, ни в части эффективной организации вычислений на современных компьютерах. При этом немаловажным препятствием являются не вполне оправданные стереотипы, сложившиеся в обработке изображений. Так в современных учебниках по распознаванию изображений утверждается, что критерий СКО «плохо

согласуется с критерием субъективного восприятия» ([3]), что, как правило, наблюдается при сравнении по СКО двух различных изображений [4,5] и не наблюдается, если изображение, согласно кластерному анализу [1, 2], сравнивается со своим достаточно близким кусочно-постоянным приближением. Другим стереотипом является утверждение, что задача минимизации СКО является NP-полной, на основании которого делается вывод, что эффективного решения, «не зависящего от структуры данных и специфики решаемой задачи, не может существовать даже теоретически [6]» ([7]). При этом обычно не принимается во внимание, что подобные соображения малосодержательны, если речь не идет о точном решении оптимизационной задачи.

При недостатке готового ПО для кластеризации пикселей опыт программистов по оптимизации кусочно-постоянных приближений обычно ограничивается применением малоэффективного метода K-средних, который не следует из условия минимизации СКО и, в общем случае, завершается до достижения возможного минимума СКО. При этом возможности иерархической кластеризации пикселей часто ограничиваются применением аппарата стандартных деревьев в виде дендрограмм, который приводит к ресурсозатратным вычислениям.

Обсуждаемая модель иерархической кластеризации пикселей позволяет преодолеть отмеченные недостатки.

Модель детектирования объектов

В отличие от большинства технических решений, разрабатываемых в России для текущих прикладных задач, обсуждаемая модель иерархической кластеризации пикселей предназначается для обработки изображений, ни содержание которых, ни целевые «объекты интереса» заранее не ограничивается. Подобный подход к распознаванию изображений применяется преимущественно за рубежом при детектировании на изображении «наиболее заметных областей» (salient region detection) [8,9]. При этом для изображения строится, так называемая, карта заметных областей, которая обычно преобразуется в черно-белую маску «объект-фон» пороговым преобразованием. В модели иерархической кластеризации пикселей ставится более конкретная задача упорядочения всех рассматриваемых множеств пикселей по единому параметру.

Для входного изображения в модели строится, запоминается и, при необходимости, преобразуется бинарная иерархия кластеров которая содержит N приближений изображения. Приближения содержат всего $2N-1$ различных кластеров пикселей. Из них N кластеров неделимы, т к. состоят из отдельных пикселей, а для каждого из $N-1$ остальных кластеров поддерживается операция разделения надвое, при которой запоминается и, по мере необходимости, восстанавливается пара кластеров пикселей 1 и 2, слиянием которых данный кластер 3 получен: $3 \to 1,2 : 3 = 1 \cup 2$.

Под *детектированием* (фильтрацией) объектов подразумевается последовательный выбор кластеров пикселей при условии однозначного отнесения каждой точки изображения либо к объекту, либо к фону. Под *иерархическим детектированием* (иерархической фильтрацией) объектов подразумевается последовательное проявление или стирание объектов При проявлении объектов фон задается определенным набором координат и сокращается по с увеличением числа объектов, проявляемых в порядке уменьшения значений некоторого атрибута, который имеет смысл количественной меры.

Признак F считается *количественной мерой*, если он не увеличивается, когда кластер $1 \cup 2$ делится на два вложенных кластера 1 и 2:

$$F(1 \cup 2) \geq F(1),\ F(1 \cup 2) \geq F(2) \qquad (1)$$

Характерным примером признака типа количественной меры является число n пикселей в кластере, для которого свойства (1) очевидны.

Следует отметить, что на начальном этапе кластеризации пикселей, предназначенном для последующей фильтрации объектов, возникает проблема упорядочения кластеров пикселей (цветов) по некоторому признаку, называемому в [8, 9] «заметностью» («saliency»), «контрастностью» или, в терминологии [10], «сложностью» фрагментов изображения. В обсуждаемой модели детектирования объектов искомый признак естественным образом получается при ап-

проксимации изображения квазиоптимальными приближениями, которые по ошибке аппроксимации $E = 3N\sigma^2$ близки к оптимальным приближениям изображения с минимальной ошибкой. Хотя практически вычислить последовательность оптимальных приближений цветового изображения с числом кластеров от 1 до N, в общем случае, не представляется возможным, тем не менее, известно, что эта последовательность описывается монотонно возрастающей последовательностью $\Delta E_2 \le \Delta E_3 \le ... \le \Delta E_{N-1} \le 0$ неположительных приращений E или выпуклой последовательностью самих значений E:

$$E_g \le \frac{E_{g-1} + E_{g+1}}{2}, \quad g = 2, 3, ..., N-1. \tag{2}$$

В общем случае последовательность оптимальных приближений не является иерархической, в чем нетрудно убедиться на примере изображения с линейным возрастанием яркости по одной из координат. Для того, чтобы избежать принципиальных вычислительных трудностей в развиваемой модели детектирования объектов решается задача аппроксимации неирархической последовательности оптимальных приближений иерархией приближений с числом g кластеров пикселей от 1 до N. При этом требуется, чтобы аппроксимирующая иерархия приближений в зависимости от числа кластеров g описывалась выпуклой последовательностью ошибок аппроксимации E_g. В этом случае, благодаря условию выпуклости гарантируются ограниченные значения ошибки аппроксимации E_g, которые не превышают определенного порога:

$$E_g \le E_1\left(1 - \frac{g}{N}\right), \tag{3}$$

где E_1 ошибка аппроксимации изображения одинаковыми пикселями средней яркости.

Аппроксимация изображения квазиоптимальными приближениями иллюстрируется рис. 1.

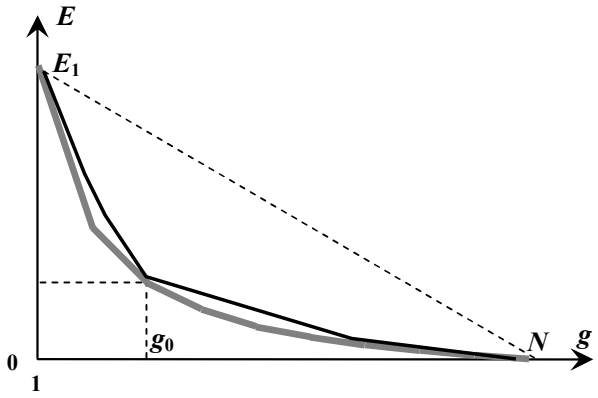


Рис. 1. Аппроксимация изображения квазиоптимальными приближениями.

На рис. 1 показана зависимость ошибки аппроксимации E от числа кластеров g. Серая кривая описывает оптимальные приближения изображения. Сплошная черная кривая описывает иерархию квазиоптимальных приближений. Обе кривые выпуклые и имеют общее начало и конец. При некотором значении $g = g_0$ кривая для квазиоптимальных приближений максимально приближается или соприкасается с кривой для оптимальных приближений, что достигается за счет минимизации ошибки аппроксимации для заданного значения $g_0 = 2,3,...$ При этом g_0 является настроечным параметром, который предусматривается в модели детектирования объектов. Предполагается, что при $g = g_0$ изображение разделяется на целостные объекты, части которых не встречаются при $g < g_0$, тогда как объединения объектов не встречаются при $g > g_0$. Эти соображения позволяют выбрать параметр g_0 для наилучшего обнаружения конкретных «объектов интереса».

Формально, требования к минимизации ошибки аппроксимации E изображения посредством приближения с g_0 кластерами пикселей выражаются следующими двумя утверждениями:

- Итеративное разделение каждого кластера надвое в зависимости от количества кластеров описывается выпуклой последовательностью значений E.
- Ошибка аппроксимации E для заданного количества g_0 кластеров пикселей (цветов в приближении изображения) не может быть уменьшена путем разделения одного кластера надвое и слияния двух других кластеров.
- Для заданного количества g_0 кластеров пикселей ошибка аппроксимации не может быть уменьшена путем реклассификации (переноса) той или иной предусмотренной части из одного кластера пикселей в другой кластер пикселей.

Здесь первое условие обеспечивается итеративной кластеризацией пикселей по методу Уорда [1, 2, 11]. Второе условие обеспечивается применением версии SI-метода (Segmentation Improvement-метода) [12] для снижения ошибки аппроксимации E. Третье условие обеспечивается минимизацией E и σ методом «K-meanless» [13, 14].

Оригинальная версия метода Уорда [11] обеспечивает выполнение первого и второго условий, но требует чрезмерно много времени. В ускоренном варианте первые два условия обеспечиваются поочередным применением SI-метода в сочетании с обработкой методом Уорда изображения *по частям*, т.е. по кластерам пикселей, обрабатываемым как самостоятельные изображения. Если обсуждаемые условия выполняются, то иерархия приближений с числом кластеров g от g_0 до N методом Уорда достраивается до полной иерархии приближений с g от 1 до N, описываемой выпуклой последовательностью $E_1, E_2, ..., E_N$ значений ошибки аппроксимации E. При этом благодаря второму условию не возникает нарушения выпуклости при g_0 кластерах. Таким путем гарантируется исправление грубых начальных приближений изображения. Третье условие при выполненных первых двух обеспечивает дальнейшую минимизацию E_g при $g = g_0$ кластерах.

В целом, обсуждаемые три условия обеспечивают аппроксимацию оптимальных приближений квазиоптимальными приближениями, и аппроксимацию выпуклой последовательности оптимальных значений $E_g^{\min}$ такой же выпуклой последовательностью значений E_g, с минимизированной разницей $E_g - E_g^{\min} \approx 0$ при $g = g_0$ кластерах.

Все иерархически структурированные кластеры пикселей упорядочиваются по абсолютной величине приращения ошибки аппроксимации $|\Delta E|$, сопровождающей разделение данного кластера надвое, которая для неделимых пикселей полагается равной 0 и является количественной мерой (1):

$$\Delta E(1 \cup 2) \geq \Delta E(1), \ \Delta E(1 \cup 2) \geq \Delta E(2). \qquad (4)$$

Детектирование объектов сводится к фильтрации кластеров, которые выбираются из общего числа $2N-1$ иерархически структурированных кластеров пикселей по заданному значению порога $T \geq 0$. При этом из иерархической последовательности N разбиений отбираются неповторяющиеся разбиения $g_1, g_2, g_3 \ldots$, содержащие как кластеры со значением $|\Delta E| \leq T$, так и кластеры со значением $|\Delta E| \geq T$:

$$g_1, g_2, g_3 \ldots: \quad |\Delta E|_{\max}(g) \geq T \geq \min |\Delta E|_{\min}(g), \qquad (5)$$

В этом случае, пиксели кластеров со значением $|\Delta E| \leq T$ относят к пикселям «объектов интереса», а кластеры со значением $|\Delta E| \geq T$ относят к фону, или наоборот. Эквивалентным образом можно выделять объекты по количеству пикселей в кластерах и получать последовательности разбиений $g_1^n, g_2^n, g_3^n \ldots$, содержащие кластеры с числом пикселей большим и меньшим заданного порога T_n:

$$g_1^n, g_2^n, g_3^n \ldots: \quad n_{\max}(g) \geq T_n \geq n_{\min}(g), \tag{6}$$

Физический смысл формул (5)-(6) состоит в том, что они описывают получение иерархии объединений одного, двух и более объектов по единственному значению параметра количественной меры $|\Delta E|$ или n.

Аналитическое описание

Апостериори любую выпуклую последовательность $E_{g=1}$, $E_{g=2}$, ..., $E_{g=N}$, можно получить укрупнением пикселей методом Уорда [11], в котором вначале каждый пиксель составляет самостоятельный кластер. Затем на каждом шаге сливается пара кластеров i, j, отвечающая минимальному приращению ошибки аппроксимации $\Delta E_{merge}(i, j)$:

$$i, j \to i \cup j: \quad i, j = \underset{i, j=0,2,\ldots g-1}{\arg\min} \; \Delta E_{merge}(i, j), \tag{7}$$

где число кластеров снижается от N до 1, а приращение ошибки аппроксимации $\Delta E_{merge}(i, j)$ выражается через число писелей n_i, n_j и трехкомпонентные средние значения I_i, I_j пикселей внутри кластеров i, j в виде:

$$\Delta E_{merge}(i, j) = -\Delta E(i \cup j) = \frac{n_i n_j}{n_i + n_j} \left\| I_i - I_j \right\|^2 \geq 0, \tag{8}$$

где $\Delta E(i \cup j) \leq 0$ неположительная величина приращения ошибки аппроксимации при разделении надвое кластера $i \cup j$.

Формула (8) дает явное выражение для величины $|\Delta E(i \cup j)| = \Delta E_{merge}(i, j)$, являющейся количественной мерой яркостной неоднородности или сложности фрагмента изображения, который составляют пиксели данного кластера $i \cup j$.

В силу выраженной повторяемости пар кластеров i, j, отвечающих минимальным значениям $\Delta E_{merge}(i, j)$, на начальных шагах слияния кластеров, результаты итеративного слияния кластеров методом Уорда зависят от выбора порядка слияния и описывается множеством переплетающихся выпуклых последовательностей значений ошибки аппроксимации $E_{g=1}$, $E_{g=2}$, ..., $E_{g=N}$ [15], как показано на рис.2

Выбор иерархии и отвечающей ей кривой $E_{g=1}$, $E_{g=2}$, ..., $E_{g=N}$, в точке $g = g_0$ приближающейся к кривой $E_{g=1}^{\min}$, $E_{g=2}^{\min}$, ..., $E_{g=N}^{\min}$, для оптимальных значений, обеспечивается методом K-средних без средних, в котором для иерархии частей j кластеров i минимизируется (если достигается) отрицательное приращение $\Delta E_{\text{correct}}(i, j, k)$ ошибки аппроксимации:

$$i, j, k = \underset{i, j=1,2,\ldots,g_0,\, k=1,\ldots,2n_i-1}{\arg\min} \Delta E_{correct}(i, j, k) < 0, \tag{9}$$

где приращение $\Delta E_{correct}(i, j, k)$ ошибки аппроксимации при реклассификации k –й части кластера i в кластер j выражается в виде:

$$\Delta E_{correct}(i, j, k) = \frac{n_j n_k}{n_j + n_k} \left\| I_j - I_k \right\|^2 - \frac{n_i n_k}{n_i - n_k} \left\| I_i - I_k \right\|^2, \tag{10}$$

где n_i, n_j, $n_k < n_i$ и I_i, I_j, I_k — значения числа пикселей и трехмерных средних яркостей в кластерах i, j, k, соответственно.

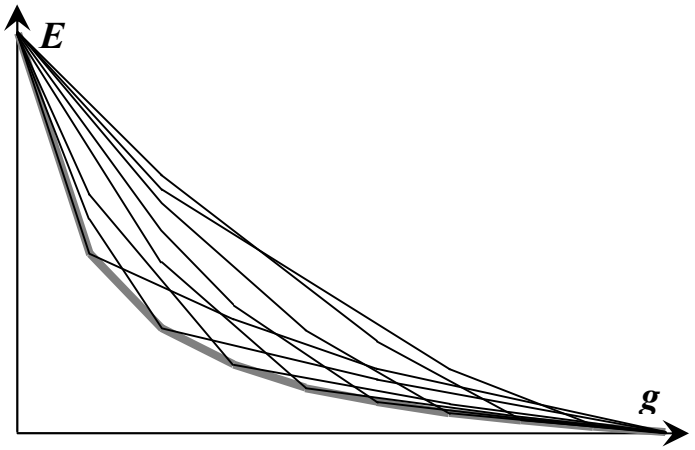


Рис. 2. Неоднозначность аппроксимация изображения квазиоптимальными приближениями.

Выписанная формула следует из условия минимизации E или σ. Если в ней опустить дробные коэффициенты, учитывающие количества пикселей n_i, n_j, n_k в кластерах i, j, k, то получится критерий минимизации метода *K*-средних. При расчетах с помощью линейки, циркуля и арифмометра метод *K*-средних позволяет избежать операций с чрезмерно большими числами и трудоемкой операции извлечения квадратного корня на арифмометре. Но в приложении к кластеризации пикселей этот метод приводит к большому числу ложных решений, главным образом из-за нарушения порядка слияния/разделения кластеров, выбираемого из условия минимизации приращения ошибки аппроксимации ΔE на каждом шаге формирования бинарной иерархии приближений изображения.

Заключение

Таким образом, в докладе описаны основные положения модели квазиоптимальной кластеризации пикселей изображения, и предложен способ порогового детектирования иерархии объединений объектов, в котором для детектирования объектов оказывается достаточным указание единственного порога предельного значения признака типа количественной меры.

Помимо представленной наглядной модели квазиоптимальной кластеризации пикселей изображения, для ее реализации разработана сетевая структура данных, необходимая для скоростных вычислений. В этой структуре данных вычисления выполняются в терминах деревьев (ациклических графов) и циклов (циклических графов). Причем, в упомянутой структуре данных используются не традиционные, а динамические деревья Слейтора-Тарьяна [16-18], которые вместе с циклами образуют динамическую сеть и обеспечивают построение, запоминание, а также преобразование миллионов кусочно-постоянных приближений изображения в RAM. Особенности скоростных вычислений также будут описаны в докладе.

Реализованный в терминах динамических деревьев аппарат детектирования объектов на основе генерации и оптимизации кусочно-постоянных приближений изображения по ошибке аппроксимации E или среднеквадратичному отклонению σ (СКО), по всей видимости, дополняет разработки зарубежных ученых по созданию инструментария обработки изображений в рамках вероятностного подхода.

ЛИТЕРАТУРА

1. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерности. М.: Финансы и статистика, 1989. 607 с.
2. Мандель И.Д. Кластерный анализ. М.: Финансы и статистика. 1988. 176 с.
3. Сойфер В.А. (Ред.) Методы компьютерной обработки изображений. М.: Физматлит, 2001. 784 с.
4. Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P. Image quality assessment: From error visibility to structural similarity // IEEE Transactions on Image Processing, Apr. 2004. Vol. 13, № 4, pp. 600-612.
5. Blasch E., Li X., Chen G., Li W. Image Quality Assessment for Performance Evaluation of Image Fusion // Int. IEEE Conf. on Information Fusion, 2008. pp. 583-588.

6. Xu R., Wunch D.C. Clustering . N.Y.: John Wiley&Sons, 2009. 358 p.
7. Пестунов И.А., Синявский Ю.Н. Алгоритмы кластеризации в задачах сегментации спутниковых изображений // Вестник КемГУ. 2012. №2 С.110-125.
8. Achanta R., Hemami S., Estrada F., Susstrunk S. Frequency-tuned salient region detection // Computer vision and pattern recognition (CVPR), IEEE conference, 2009. pp. 1597-1604.
9. Cheng M.M., Mitra N.J., Huang X., Torr P.H., Hu S.M. Global contrast based salient region detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2015. Т. 37. №. 3. pp. 69-582.
10. Чочиа П.А. Теория и методы обработки видеоинформации на основе двухмасштабной модели изображения / Дис. докт. технич. наук. М.:ИППИ РАН, 2016. 302 с.
11. Ward J.H., Jr. Hierarchical grouping to optimize an objective function J, Am. Stat. Assoc. vol. 58, Issue 301, 1963. pp. 236-244.
12. Харинов М.В., Ханыков И.Г. Оптимизация кусочно-постоянного приближения сегментированного изображения. // Труды СПИИРАН. 2015. Вып. 3(40). С. 183–202.
13. S.D. Dvoenko «Meanless k-means as k-meanless clustering with the bi-partial approach,» Proc. of the 12th International Conference on Pattern Recognition and Information Processing (PRIP'2014), Minsk, Belarus, 2014. pp. 50–54.
14. Kharinov M.V. Reclassification formula that provides to surpass K–means method /arXiv preprint, arXiv:1209.6204, 28 Sep 2012. 10 p.
15. Харинов М.В. Основы модели квазиоптимальных приближений изображения // Вестник Бурятского государственного университета, Математика и информатика / №1, Улан-Удэ: Изд-во Бурятского госуниверситета, 2016. — С. 60-72. — ISSN 2304-5728.
16. Kharinov M.V. Pixel Clustering for Color Image Segmentation // Programming and Computer Software, vol. 41, No 5, 2015, pp. 258–266.
17. Sleator D.D., Tarjan R.E. Self–Adjusting Binary Search Trees // Journal of the ACM. vol. 32 No 3, 1985. pp. 652–686.
18. Nock R., Nielsen F. Statistical Region Merging // IEEE Trans. Pattern Anal. Mach. Intell. , vol. 26 No 11, 2004.pp. 1452–1458.

M. Kharinov (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences)

Modernization of information technologiesin the field of image processing

In the report the problem of localization and detection of objects in a color image is formalized. The conventional stereotypes in the field of processing digital images, which hinder the active utilization of classical cluster analysis, are discussed. To characterize the presence of objects in the image a "measure of the intensity non-uniformity of the image areas" is introduced, which is estimated as the absolute value of the increment of the total square error when dividing of pixel clusters into two. A method for detecting of hierarchy of object associations is proposed.

ПОСТРОЕНИЕ МНОГОУРОВНЕВОЙ СИСТЕМЫ РАСПОЗНАВАНИЯ ДЛЯ ДИАГНОСТИКИ БОЛЕЗНИ ПАРКИНСОНА

О. В. Жвалевский
(*СПИИРАН, С.-Петербург, (812) 328-54-11, ozh@spiiras.ru*)

Аннотация

В докладе рассматривается задача построения системы распознавания, предназначенной для диагностики болезни Паркинсона. Основными объектами в этой системе выступают временные ряды (тензотреморограммы) и их описания в виде наборов потенциально информативных признаков и/или структурных описаний. В докладе подробным образом описывается структура экспериментальных данных (объекты, участвующие в распознавании) и приводятся допустимые классификации объектов. Делается вывод о необходимости строить многоуровневую (иерархическую) систему распознавания. В докладе описываются способы и подходы к построению многоуровневой системы распознавания. Кроме того, в докладе показывается роль в построении многоуровневой системы распознавания методов структурного анализа временных рядов, методов машинного обучения и методов математического моделирования.

Ключевые слова: Болезнь Паркинсона, тензотреморограмма, функциональное состояние, система распознавания, машинное обучение

Введение

Автоматизация диагностики какого-либо заболевания заключается в построении системы распознавания того или иного вида. Объектами в такой системе распознавания выступают физиологические сигналы — временные ряды значений физиологических показателей, содержащих информацию о функционировании определённой подсистемы организма человека [1]. Диагностика болезни Паркинсона [2] предполагает исследование многокольцевой системы построения движений по Н.А.Бернштейну [3] в норме и при различных видах патологии. Непосредственную информацию о состоянии системы построения движений предоставляет тремор рук. В частности, С.П.Романовым был предложен изометрический способ регистрации тремора [4,5], который предполагает непосредственную регистрацию усилия, оказываемого руками человека на чувствительную платформу (по одной — на каждую руку испытуемого); при этом, испытуемый следит за уровнем удерживаемого им усилия. Другими словами, здесь используется биологическая обратная связь. Тремор — это компонент, который выделяется из *произвольных колебаний*, отражающих уровень удерживаемого испытуемым усилия, и описывает *непроизвольные колебания*, которые совершаются под непосредственным управлением системы построения движений. Результаты математической обработки тензотреморограмм (ТТГ), получаемых описанным выше способом, свидетельствуют о том, что эти сигналы содержат точную и достоверную информацию о системе построения движений по Н.А.Бернштейну [6,7].

Несмотря на полученные ранее результаты, требуется получить достаточно полное и точное представление о природе анализируемых временных рядов. Во-первых, указанные выше результаты получены для весьма ограниченных выборок экспериментальных данных. В то же время, необходимо гарантировать воспроизводимость результатов. Во-вторых, указанные выше результаты получены при помощи весьма ограниченного набора подходов к математическому анализу временных рядов, хотя, некоторые из этих подходов, например, мультифрактальный анализ временных рядов [8], основанный на методе максимумов модулей вейвлет-преобразования [9], и представляются особенно эффективными. В то же время, существуют и другие подходы к анализу временных рядов, которые, вполне вероятно, позволят обнаружить новые, ранее незамеченные особенности ТТГ, например, определённые закономерности, связанные с применением биологической обратной связи, предопределяющей особо «нестационарный» характер ТТГ. Кроме того, рассмотрение различных подходов к анализу ТТГ потребует сравнительного анализа результатов, полученных при помощи различных методов. В-третьих, итогом всех исследований должна стать система поддержки принятия врачебных решений. В рамках этой системы необходимо реализовать оптимальное сочетание различных подходов к анализу ТТГ и предложить такую методику обработки экспериментальных данных,

которая позволит предоставить пользователю результаты этой обработки в простой и наглядной форме.

В работе рассматриваются различные аспекты, связанные с построением системы распознавания, предназначенной для диагностики болезни Паркинсона.

Основные объекты

В нашем распоряжении имеются временные ряды двух родов: временные ряды *первого рода* – это сигналы, описывающие произвольные колебания конечностей, которые регистрируются в ходе измерительного эксперимента непосредственно; временные ряды *второго рода* — это сигналы, описывающие непроизвольные колебания конечностей, которые выделяются из произвольных колебаний на аппаратном уровне (непосредственно перед аналого-цифровым преобразованием). Сигналы второго рода, связанные с тремором, называются, как уже было сказано выше, *тензотреморограммами*; в то же время, сигналы первого рода, описывающие уровень удерживаемого испытуемым усилия, следует условно называть, по аналогии с ТТГ, *тензофорсограммами* (ТФГ). ТФГ являются, таким образом, исходными или «сырыми» данными, в то время как ТТГ являются *вторичными* или *производными* данными.

В действительности, все полученные ранее результаты, получены при обработке именно ТТГ. Следовательно, первая задача, которая возникает при обработке экспериментальных данных, это сравнительный анализ ТТГ и ТФГ. В частности, необходимо установить, какими уникальными свойствами обладают разнородные сигналы, и как именно можно использовать информацию, извлекаемую из ТФГ, например, для обработки ТТГ. С одной стороны, проектируемая система распознавания может быть основана на математической обработке только однородных сигналов, например, ТТГ: если основная информация о системе построения движений находится в ТТГ, то использовать результаты математической обработки ТФГ нецелесообразно. С другой стороны, совместная обработка разнородных данных может дать определённые преимущества при построении системы распознавания. Если удастся установить связь между тем, как именно происходит удержание одного и того же уровня усилия различными испытуемыми, то эту связь можно использовать для более тонкого сравнительного анализа соответствующих ТТГ.

Для того, чтобы получить наиболее полную и наиболее точную информацию о системе построения движений, С.П.Романов предложил методику проведения измерительного эксперимента, составленного из четырёх различных измерительных сеансов, с каждым из которых связаны определённые условия проведения измерений (см. табл. 1).

Таблица 1

Типы измерительных сеансов

№	Постановка рук	Уровень удерживаемого усилия
1	С опорой на запястье	Минимальное
2	С опорой на запястье	Максимальное
3	С опорой на вытянутые руки	Минимальное
4	С опорой на вытянутые руки	Максимальное

Эта методика позволяет получить данные, соответствующие различным состояниям и режимам функционирования системы построения движений: сначала используется регистрация тремора с опорой на запястье (в этом случае система построения движений оказывается задействованной в наименьшей степени), затем используется регистрация тремора для вытянутых рук (в этом случае система построения движений оказывается задействованной в наибольшей степени); каждый способ регистрации используется дважды — при минимально возможном уровне удерживаемого испытуемым усилия (в этом случае система построения движений находится в минимально активном режиме функционирования) и при максимально возможном уровне удерживаемого испытуемым усилия (в этом случае система построения движений находится в максимально активном режиме функционирования). Таким образом, каждому измерительному сеансу соответствует свой тип экспериментальных данных или, что то же самое, свой тип ТТГ (и ТФГ).

Построение системы распознавания, основанной на математическом анализе разнотипных данных, может быть осуществлено при помощи трёх различных подходов. Первый подход заключается в сравнительном анализе ТТГ различных типов и выборе того типа ТТГ, который лучше всего подходит для построения системы распознавания. Поскольку, при построении каждой отдельной системы распознавания учитываются только свойства временных рядов соответствующего типа, такой подход обладает очевидными ограничениями, но в качестве первичного «разведочного» анализа имеющихся экспериментальных данных этот подход может быть вполне пригодным. Второй подход заключается в построении иерархической системы коллективного распознавания, на нижнем уровне которой располагаются частные системы распознавания, построенные для временных рядов определённого типа, а на верхнем уровне располагается решающие правило, использующее в качестве входных данных результаты работы частных систем распознавания. В самом простом варианте реализации данного подхода, используется какая-либо процедура голосования. В наиболее сложном варианте реализации данного подхода решающие правила частных подсистем распознавания оказываются в зависимости от решающих правил вышележащего уровня. Наконец, третий подход заключается в применении методов комплексной обработки разнотипных экспериментальных данных, когда для построения решающих правил используются все имеющиеся экспериментальные данные.

Основная особенность регистрируемых сигналов (временных рядов) — это то, что каждый род сигналов представлен двумя одновременными друг другу временными рядами: один из них соответствует усилию или тремору левой руки, а другой — усилию или тремору правой руки. Это открывает дополнительные возможности для построения системы распознавания. Во-первых, необходимо провести сравнительный анализ временных рядов различных видов (представленных в табл. 2) и выяснить, можно ли каким-нибудь образом представить два одновременных временных ряда в виде единого временного ряда, содержащего наиболее важную информацию о характере удержания испытуемым усилия. Это, очевидно, позволит существенно упростить построение системы распознавания. При этом, также очевидно, что основное внимание должно быть уделено, при таком подходе, именно «сырым» данным, описывающим усилие, но, также возможен и анализ тремора (по той же схеме). Во-вторых, необходимо провести сравнительный анализ одновременных однородных временных рядов и выяснить, в каких случаях «разносторонние» временные ряды наиболее «близки» друг к другу, а в каких случаях «разносторонние» временные ряды наиболее «удалены» друг от друга. Это может иметь значение для определения ранних признаков болезни Паркинсона, когда начальные проявления заболевания обнаруживаются только на одной стороне. Предполагаемая здесь мера различия (между одновременными временными рядами) должна в максимально формализованном виде описывать характер удержания испытуемыми задаваемого уровня усилия и являться, таким образом, важным информативным признаком, дополняющим те признаки, которые вычисляются при помощи стандартных методов математического анализа временных рядов. В-третьих, необходимо использовать методы комплексной обработки разнотипных экспериментальных данных и выяснить, можно ли использовать результаты обработки временных рядов одного вида (например, усилия) для обработки временных рядов другого вида (в данном случае, тремора).

Таблица 2

Виды временных рядов

№	Временной ряд
1	Тремор левой руки
2	Тремор правой руки
3	Усилие левой руки
4	Усилие правой руки

В итоге, в нашем распоряжении имеются разнотипные экспериментальные данные, представленные временными рядами различных видов и типов, а также всевозможные потенциально информативные признаки, которые можно получить, обрабатывая имеющиеся временные ряды при помощи различных методов математического анализа временных рядов. В общей сложности, имеется 16 временных рядов различных типов. При этом, временные ряды одного типа связаны определёнными зависимостями с временными рядами другого типа. Эти зависи-

мости включают номинальную зависимость (когда рассматриваются два одновременных друг другу и одноимённых временных ряда), функциональную зависимость (когда один временной ряд получается из другого временного ряда при помощи того или иного преобразования) и структурную зависимость (когда рассматриваются однородные временные ряды, которые получены при проведении различных измерительных сеансов). Указанные зависимости позволяют связывать воедино разнотипные и разнородные потенциально информативные признаки и могут быть использованы для более эффективного построения решающих правил.

Классификация объектов

Система распознавания, предназначенная для диагностики болезни Паркинсона, в самом первом и грубом приближении представляет собою систему решающих правил, позволяющих относить испытуемых к одному из четырёх классов, перечисленных в табл. 3. Изначально предполагается, что каждый испытуемый находится в определённом функциональном состоянии (ФС), которое соответствует определённому характеру функционирования системы построения движений. Соответственно, если система построения движений функционирует нормально, то следует говорить о том, что испытуемый находится в ФС «в норме»; если, однако, система построения движений функционирует не так, как следует, то следует говорить о том, что испытуемый находится в ФС «в патологии», при этом, следует указывать, какой именно вид патологии имеется у данного испытуемого. В этом смысле, система распознавания производит оценку ФС испытуемого в слабой шкале наименований. Табл. 3, представляет собою, как раз, пример слабой шкалы наименований, которая используется при построении системы распознавания в самом первом и грубом приближении.

Таблица 3

Классы испытуемых (функциональных состояний)

№	Наименование класса (функционального состояния)
1	Практически здоровые испытуемые
2	Люди, больные болезнью Паркинсона (БП)
3	Люди, больные синдромом паркинсонизма (СП)
4	Люди, больные общей патологией, не связанной ни с БП, ни с СП

Приведённая выше слабая шкала наименований обладает рядом существенных недостатков.

Во-первых, оценка ФС осуществляется в определённый момент времени — в момент регистрации физиологических сигналов. Это означает, что с участием одного и того же испытуемого может быть произведено несколько измерительных экспериментов в различные моменты времени, и для каждого измерительного эксперимента может быть получена, вообще говоря, своя оценка ФС. Если испытуемый действительно здоров (если у испытуемого нет никакой патологии), то можно ожидать, что результаты оценки ФС, полученные в различные моменты времени, будут совпадать. Если, однако, изначально практически здоровый испытуемый заболел болезнью Паркинсона, то, со временем, оценка его ФС должна измениться. Если бы у нас была возможность определять стадию развития болезни Паркинсона, то нам следовало бы использовать для оценки ФС *порядковую* шкалу. При этом, принципиальное значение будет иметь наличие в регистрируемых физиологических сигналах ранних признаков развития заболевания, которые можно выделить из физиологических сигналов при помощи методов математического анализа временных рядов. Если моторные проявления болезни Паркинсона, обнаруживаются на позднем этапе развития заболевания, то метод треморографии не может использоваться для ранней диагностики болезни Паркинсона. В то же время, объективный анализ ТТГ призван для того, чтобы выявить основные признаки «нормы» и «патологии», и должен быть использован для построения порядковой шкалы наименований.

Во-вторых, оценка ФС людей, у которых имеется та или иная патология, существенно затруднена. В частности, люди, больные болезнью Паркинсона (или синдромом паркинсонизма), обычно, принимают определённые лекарственные препараты, предназначенные для подавления симптомов заболевания. Для того, чтобы оценить «истинное» ФС такого испытуемого, приходится временно отменять приём лекарственного препарата. В то же время, обычное ФС испы-

туемого, больного болезнью Паркинсона (или синдромом паркинсонизма), суть «изменённое» ФС. С одной стороны, наличие «изменённых» ФС затрудняет построение системы распознавания, поскольку необходимо предпринимать специальные усилия, чтобы получить данные, соответствующие патологии. С другой стороны, необходимо оценивать эффективность лекарственной терапии: если «изменённое» ФС очень «похоже» на ФС «в норме», то такую лекарственную терапию следует признать эффективной.

В-третьих, проблему вызывает само понятие ФС. На самом деле, истинное состояние системы построения движений нам неизвестно. Осуществляя регистрацию физиологических сигналов, мы вводим систему построения движений в один из возможных режимов функционирования. Имея четыре различных постановки измерительного эксперимента, мы имеем, фактически, четыре различных режима функционирования системы построения движений. Предполагая, что у практически здоровых испытуемых и у людей, у которых имеется какая-либо патология, система построения движений функционирует по-разному, мы хотим максимальным образом проявить эти различия, задавая различные режимы функционирования. При этом, оказываются важными не только индивидуальные различия, имеющиеся между физиологическими сигналами, полученными для различных классов испытуемых в одном и том же режиме функционирования системы построения движений, но и то, как именно меняются характеристики физиологических сигналов при изменении режима функционирования, например, при переходе от минимально возможного уровня удерживаемого усилия к максимально возможному, или при переходе от одной постановки рук (при опоре на запястье) к другой постановке рук (при вытянутых руках).

Представляется целесообразным описывать норму и различные виды патологии при помощи одной шкалы наименований (это может быть слабая шкала наименований, содержащая, в том числе, все возможные варианты патологий, или порядковая шкала, описывающая различные стадии развития заболевания), а различные состояния, в которых может находиться система построения движений, описывать при помощи другой шкалы наименований, которую ещё необходимо построить, пользуясь результатами математического анализа физиологических сигналов. Другими словами, речь идёт о применении подхода к построению системы распознавания, основанного на марковских моделях со скрытыми состояниями. Таким образом, имея физиологические сигналы, полученные для нормы и для различных видов патологии, необходимо реконструировать те неизвестные состояния, в которых может находиться система построения движений у каждого класса испытуемых. При этом, сама оценка ФС относится к конкретным физиологическим сигналам, полученным в определённый момент времени и для определённого испытуемого, в то время как реконструируемая структура неизвестных заранее скрытых состояний относится к норме и к различным видам патологии. Это означает, что за нормой и за различными видами патологии стоит своя марковская модель, а система распознавания осуществляет идентификацию этой модели.

Таким образом, проектируемая система распознавания должна быть иерархической. На верхнем уроне этой иерархии находятся состояния системы построения движений «в норме» и при различных видах патологий данной системы с учётом того, такие состояния являются «истинными», а какие — «изменёнными», и с учётом того, какие стадии развития может проходить болезнь Паркинсона и/или синдром паркинсонизма. На среднем уровне этой иерархии располагаются те неизвестные нам ФС, в которых может находиться система построения движений в зависимости от того, в каком режиме функционирования оказывается система построения движений в результате проведения измерительного эксперимента определённого типа. Наконец, на нижнем уровне этой иерархии располагаются задаваемые в эксперименте режимы функционирования системы построения движений, которые полностью определяются тем, каким именно образом производится измерительный эксперимент. В такой иерархии нижний и верхний слои содержат либо то, что известно заранее или может быть задано заранее (в нашем случае это набор допустимых режимов функционирования системы построения движений, определяемых списком градаций уровня удерживаемого испытуемым усилия), либо то, что в целом можно считать известным (в нашем случае это набор различных видов патологии) или требующим реконструкции (в нашем случае это набор различных стадий болезни Паркинсона или синдрома паркинсонизма); в то же время, средний слой содержит то, что неизвестно заранее и требует

обязательной реконструкции (в нашем случае это структура ФС с учётом того, какие могут быть нормальные состояния и различные виды патологии, и с учётом того, какие могут быть переходы между различными ФС). В соответствии с этим, будем условно называть каждый вариант нормы и каждый вид патологии, а также, возможно, и различные стадии развития той или иной патологии, (функциональным) *макросостоянием*, подразумевая под этим, что каждое функциональное макросостояние содержит определённое множество обыкновенных ФС системы построения движений. В этом смысле, система распознавания должна осуществлять оценку именно функциональных макросостояний, и эта оценка непосредственным образом опирается на оценку ФС системы построения движений и использует результаты этой оценки для реконструкции структуры функциональных макросостояний.

Основным источником данных для построения иерархической системы распознавания являются временные ряды — физиологические сигналы, регистрируемые в различных режимах функционирования системы построения движений. Набор допустимых режимов функционирования, размерность и конкретный состав анализируемых временных рядов, а также выбираемые методы математического анализа временных рядов предопределяют структуру экспериментальных данных. В то же время, сами временные ряды также могут обладать некоторой структурой.

Если, например, задана некоторая сегментация, и на каждом сегменте данного временного ряда при заданной сегментации вычисляется один и тот же набор потенциально информативных признаков, то мы получаем представление временного ряда в виде числовой матрицы, содержащей сжатое признаковое описание целого временного ряда. Вместо того, чтобы вычислять какие-либо интегральные показатели и использовать их в качестве потенциально информативных признаков, мы можем непосредственно использовать матричные представления и формировать сборные матрицы экспериментальных данных, каждая строка которых соответствует отдельному сегменту определённого временного ряда, а каждому временному ряду, в таком случае, будет соответствовать сразу несколько строк. Такой подход позволяет строить решающие правила, описывающие более тонкие зависимости. В частности, появляется возможность отбора только «правильных» (в смысле некоторого задаваемого заранее критерия) сегментов и использования только «правильных» сегментов для построения решающих правил. Кроме того, усиление или ослабление критерия отбора «правильных» сегментов позволяет строить гибкие (настраиваемые) решающие правила.

Более глубокий *структурный анализ временных рядов* [10] приводит к тому, что каждый временной ряд можно будет представить в виде упорядоченного набора объектов, которые, в данном случае, называются *событиями*. Каждое событие — это некоторый *типовой* фрагмент временного ряда. Основное предположение структурного анализа временных рядов заключается в том, что различным ФС системы построения движений соответствуют временные ряды, обладающей различной «событийной» структурой. Это означает, что различным ФС соответствуют различные «словарные запасы» типовых фрагментов и различные правила составления правильных «предложений» из типовых фрагментов. Такой способ представления временных рядов снова приводит нас к необходимости использовать подход, основанный на марковских моделях со скрытыми состояниями. Только, на этот раз, в качестве неизвестных («скрытых») состояний выступают типовые фрагменты анализируемых временных рядов. Таким образом, эффективность системы распознавания будет существенным образом зависеть от того, какая будет проведена классификация фрагментов анализируемых временных рядов.

Многоуровневая система распознавания

Построение системы распознавания осуществляется *снизу вверх*: сначала рассматриваются имеющихся временные ряды; затем, эти ряды обрабатываются различными методами математического анализа временных рядов, и вычисляются всевозможные потенциально информативные признаки; далее, выбираются методы распознавания и критерий качества распознавания; в самом конце, отбираются информативные признаки и строятся основанные на них решающие правила.

Существует два различных подхода к построению системы распознавания.

Первый подход заключается в разработке специализированного «сильного» метода математического анализа временных рядов, который максимальное полно и точно учитывает все особенности анализируемых временных рядов и позволяет вычислять информативные признаки наивысшего качества. В этом идеальном случае, система решающих правил оказывается наиболее простой (из всех возможных) и наиболее надёжной, а применение методов машинного обучения сводится к определению параметров выбранной относительно простой модели, используемой для построения решающих правил. В действительности, такой идеальный случай является большой редкостью, и, поэтому, приходится не только рассматривать несколько различных методов математического анализа временных рядов, но и несколько различных методов машинного обучения. В этом случае, становится актуальной проблема *переобучения* или, что то же самое, проблема *обобщающей способности* проектируемой системы решающих правил. Как правило, в реальной ситуации, объекты различных классов довольно сильно перемешаны в признаковом пространстве, что существенным образом затрудняет построение *устойчивых* решающих правил.

Второй подход к построению системы распознавания заключается в построении коллективов решающих правил или, что то же самое, ансамбля «слабых» алгоритмов, каждый из которых связан с отдельным аспектом анализируемых временных рядов или с отдельным типом связей, присутствующих между различными фрагментами экспериментальных данных. В этом случае, на первый план выходят методы *препроцессинга*, которые позволяют представить временные ряды в форме, наиболее удобной для построения решающих правил. Во-первых, такие представления должны сжатыми или компактными и, в этом смысле, более удобными для анализа. Во-вторых, такие представления должны описывать только наиболее важные для анализа свойства временных рядов. В-третьих, такие представления должны обладать свойством инвариантности по отношению к определённым классам преобразований исходных данных. Кластеризация анализируемых временных рядов, осуществляемая после применения методов препроцессинга, необходима для того, чтобы представить имеющиеся экспериментальные данные в сжатой и наглядной форме, а также для того, чтобы выбрать подходящие методы математического анализа временных рядов, позволяющие вычислять информативные признаки, наиболее полно и точно отражающие основные закономерности, содержащиеся в экспериментальных данных. Наконец, этап машинного обучения предполагает применение значительного количества различных алгоритмических моделей, основанных на одних и тех же или различных признаках, которые, по отдельности, дают относительно «слабые» результаты при распознавании, но, будучи объединёнными в ансамбль, позволяют получать достаточно «сильные» результаты при распознавании [11].

Логически завершённым вариантом реализации второго подхода являются методы глубокого обучения [12], которые основаны на многослойных нейронных сетях и, в особенности, на свёрточных нейронных сетях, каждый слой которых связан со своим уровнем абстракции: сначала рассматриваются только самые простейшие свойства анализируемых временных рядов, строятся разного рода модели с различной обобщающей способностью; затем, эти свойства обобщаются на более высоком уровне рассмотрения и строятся новые модели; построения завершаются при достижении глобального оптимума, связанного с обобщающей способностью всей системы распознавания. Успешные примеры применения многослойных нейронных сетей позволяют использовать методы глубокого обучения практически в любой области исследований. Однако, вопрос интерпретируемости получаемых при помощи методов глубокого обучения результатов, особенно если учитывать такую область как медико-биологические исследования, становится особенно актуальным. Это означает, что результатами, даже если они оцениваются как успешные, необходимо пользоваться с некоторой осторожностью. Именно поэтому, наиболее удачным представляется формирование гипотез, основанных на результатах применения методов машинного обучения, и проверка этих гипотез при помощи независимых методов.

Построение системы распознавания следует осуществлять также и *сверху вниз*. В этом случае, построение системы распознавания заключается в применении методов математического моделирования. Применение методов математического моделирования позволяет связать между собою различные уровни многоуровневой системы распознавания при помощи задания зависимостей результатов распознавания одного уровня системы от результатов распознавания

другого уровня системы. В результате всех построений, возникает разветвлённая иерархия решающих правил, алгоритмических моделей и математических моделей различных типов. Таким образом, концептуальная модель (обобщённая схема) математической обработки ТТГ представляет собою иерархию, на нижнем уровне которой находятся структурные элементы анализируемых временных рядов (события); последовательное восхождение по иерархии — это процесс последовательной идентификации моделей различных типов; а на верхнем уровне иерархии находится финальное решающее правило, выражающее собою конечный результат работы системы распознавания.

Заключение

Автоматизация диагностики болезни Паркинсона предполагает регистрацию физиологических сигналов, несущих информацию о системе построения движений, а также разработку специализированных методов математического анализа физиологических сигналов, позволяющих (методов) вычислять информативные признаки и строить основанные на них эффективные решающие правила. Наибольшие успехи в разработке «сильного» алгоритма достигаются при применении мультифрактального анализа. В частности, фрактальная размерность и сингулярный спектр могут использоваться при различении нормы и патологии. В то же время, существуют и другие подходы к анализу временных рядов, которые также могут найти применение для обработки ТТГ. Кроме того, различные методы анализа позволяют рассматривать различные аспекты анализа и исследовать различные свойства анализируемых ременных рядов. Это побуждает использовать методы комплексной обработки экспериментальных данных и строить коллективы решающих правил. Каждое отдельное правило — это «слабый» алгоритм, связанный с определённым аспектом анализа временных рядов, однако, объединение «слабых» алгоритмов способно дать результат более надёжный результат, чем единственный, но «сильный» алгоритм.

Применение методов машинного обучения для построения решающих правил, в свою очередь, обладает рядом недостатков, из которых самый главный заключается в том, что, в обычной ситуации, задача машинного обучения формулируется и решается «как есть»: данные задаются в том виде, как они получаются на выходе алгоритмов обработки, классы объектов задаются в том виде, как они изначально формируются при проведении измерительных экспериментов и т.д. и т.п. В действительности, необходимо более тщательно подходить к отбору данных для анализа и более корректно формулировать задачу машинного обучения. Наибольшую трудность здесь вызывает само понятие ФС. Подробное рассмотрение этой проблемы заставляет существенным образом изменить постановку основной задачи распознавания и ввести дополнительный уровень для описания непосредственно связанных с нормой и различными видами патологии функциональных макросостояний, что позволяет сократить количество ошибок распознавания (а, значит, и повысить в целом обобщающую способность системы распознавания), поскольку, фактически, уточняется, какие именно объекты принимают участие в построении решающих правил и к каким именно классам в действительности относятся эти объекты. Аналогичная ситуация возникает и при анализе ТТГ, когда приходится вводить дополнительный уровень описания для структурных элементов анализируемых временных рядов и рассматривать расширенную постановку задачи распознавания, учитывающую слой промежуточных объектов. В итоге, система распознавания, предназначенная для диагностики болезни Паркинсона, оказывается многоуровневой.

Для того, чтобы построить многоуровневую систему распознавания, необходимо использовать методы математического моделирования. В свою очередь, математическое моделирование предполагает построение моделей анализируемых временных рядов, описывающих, по своей сути, алгоритм порождения временных рядов, моделей функциональных макросостояний, описывающих структуру переходов между различными ФС, а также моделей, описывающих саму систему построения движений по Н.А.Бернштейну. Даже если имеется относительно простая задача машинного обучения, требуется построение всех указанных выше элементов многоуровневой системы распознавания, поскольку только в этой ситуации появляется возможность получить надёжные и интерпретируемые результаты. Таким образом, построение системы рас-

познавания, осуществляемое, обычно, снизу вверх, будет подкрепляться результатами построения системы распознавания, осуществляемого сверху вниз.

ЛИТЕРАТУРА

1. Бернштейн Н.А. О построении движений. М.: Медгиз, 1947. 254 с.
2. Крыжановский Г.Н., Карабань И.Н., Магаева С.В. Болезнь Паркинсона (Этиология, патогенез, клиника, диагностика, лечение, профилактика). Москва: Медицина, 2002. 334 с.
3. Бернштейн Н.А. О построении движений. М.: Медгиз, 1947. 254 с.
4. Романов С.П., Якимовский А.Ф., Пчелин М.Г. Метод тензометрии для количественной оценки тремора // Физиологический журнал им. И.М. Сеченова. 1996, т. 82, № 2. С. 118–123.
5. Романов С.П., Пчелин М.Г., Якимовский А.Ф. Характеристики изометрически регистрируемого тремора при поражении экстрапирамидной системы // Российский физиологический журнал им. И.М. Сеченова. 1997, т. 83, № 3. С. 133–139.
6. Дик О.Е., Романов С.П., Ноздрачев А.Д. Энергетические и фрактальные характеристики физиологического и патологического тремора руки человека // Физиология человека, 2010, том 36, № 2, с. 92–100.
7. Дик О.Е., Ноздрачев А.Д. Нелинейная динамика непроизвольных колебаний руки человека при двигательной патологии // Физиология человека, 2015, том 41, № 2, с. 53–59.
8. Павлов А.Н., Анищенко В.С. Мультифрактальный анализ сложных сигналов // Успехи физических наук. 2007. Т. 177. С. 859.
9. Muzy J.F., Bacry E., Arneodo A. Multifractal formalism for fractal signals: the structure-function approach versus the wavelet-transform modulus-maxima method // Phys. Rev. E. 1993. V. 47. P. 875.
10. Моттль В.В., Мучник И.Б. Детерминистские модели и методы распознавания образов на оси времени. I. Основные модели и разрешимость задачи распознавания // Автоматика и телемеханика. 1991. № 3. С 120–132.
11. Воронцов К.В. Обзор современных исследований по проблеме качества обучения алгоритмов // Таврический вестник информатики и математики. — 2004.
12. Николенко С.И., Кадурин А.А., Архангельская Е.О. Глубокое обучение. Погружение в мир нейронных сетей. СПб.: Питер, 2018. 480 с.

O.V. Zhvalevsky (Federal State Institution of Science St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg)

System recognition development for Parkinson's disease diagnostics

The paper is concern with the problem of system recognition development for Parkinson's disease diagnostics. The main objects are time series (tenzotremorogramms) and it features representations. Experimental data structure and the set of objects classifications are described. The hierarchy system recognition is proposed for solving main problem. The paper deals with methods and approaches for development of hierarchy system recognition. Moreover, a role of methods for time series structural analysis, methods for machine learning and methods for mathematical modeling in development of hierarchy system recognition.

ПРИМЕНЕНИЕ МЕТОДОВ ПРЕПРОЦЕССИНГА ВРЕМЕННЫХ РЯДОВ ДЛЯ ОБРАБОТКИ ТЕНЗОТРЕМОРОГРАММ

О. В. Жвалевский
(*СПИИРАН, С.-Петербург, ozh@spiiras.ru*)

Аннотация

В докладе рассматриваются задача предварительной обработки тензотреморограмм — физиологических сигналов, возникающих при диагностике болезни Паркинсона. Этап предварительной обработки тензотреморограмм непосредственно предшествует машинному обучению. В работе в самом общем виде рассматривается задача препроцессинга временных рядов, даётся представление об основных видах препроцессинга. Так же в докладе рассматривается подход к описанию алгоритмов препроцессинга временных рядов, основанный на сегментации.

Ключевые слова: временной ряд, методы препроцессинга, спектр, инвариантное представление, сегментация

Введение

Изометрический способ регистрации тремора, предложенный С.П.Романовым, предполагает непосредственную регистрацию усилия, оказываемого руками человека на чувствительную платформу [1,2]. Фактически регистрируются *произвольные* колебания, которые совершаются конечностями. Эти колебания описывают уровень удерживаемого испытуемым усилия. Тремор — это компонент, который представляет собою *непроизвольные* колебания. Эти колебания выделяется из произвольных колебаний на аппаратном уровне при помощи интегрирующего преобразования. Сигналы, описывающие непроизвольные колебания носят название *тензотреморограммамм* (ТТГ). Изометрический способ регистрации тремора исключает движения отдельных звеньев локомоторного аппарата, что обеспечивает получение объективных данных о системе построения движений по Н.А.Бернштейну [3]. В результате проведения измерительных экспериментов, исследователь получает в своё распоряжение весьма разнообразные данные о системе построения движений. Мультифрактальный анализ [4] ТТГ показывает, что фрактальная размерность является хорошим информативным признаком [5], а сингулярный спектр хорошо описывает основные структурные особенности функционирования системы построения движений как в норме, так при патологии [6] (например, в случае болезни Паркинсона). Это означает, что для обработки ТТГ может быть применён довольно широкий набор методов математического анализа нестационарных процессов [7]. Это предполагает не только сравнительный анализ результатов, полученных при помощи различных методов, но и разработку специализированных алгоритмов обработки временных рядов, наилучшим образом учитывающих основные особенности ТТГ. Таким образом, возникает задача выбора подходящих методов математического анализа ТТГ. Эта задача решается при помощи методов машинного обучения, для применения которых формируются матрицы экспериментальных данных типа «объект-свойство» или задаются сведения относительно степени сходства/различия анализируемых временных рядов.

Особое место среди методов обработки временных рядов занимают методы предварительной обработки временных рядов или, как они ещё называются, методы *препроцессинга* временных рядов. Применение методов препроцессинга непосредственно предшествует применению методов машинного обучения. Препроцессинг временных рядов — это необходимый этап обработки экспериментальных данных, от результатов которого существенным образом зависит эффективность построения решающих правил. В соответствии с этим, представляется целесообразным, рассмотреть различные подходы к предварительной обработке временных рядов, разработать специализированные алгоритмы, реализующие эти подходы и применить эти методы для предварительной обработки ТТГ.

В работе в самом общем виде рассматривается задача препроцессинга временных рядов, даётся представление об основных видах препроцессинга. Также в работе рассматривается подход к описанию алгоритмов препроцессинга временных рядов, основанный на сегментации.

Постановка задачи

Методы препроцессинга временных рядов — это методы, которые позволяют перейти от исходного представления анализируемых временных рядов в виде упорядоченной во времени последовательности отсчётов значений наблюдаемой величины к такому представлению, которое более удобно для анализа временных рядов, чем исходное.

Во-первых, такое представление должно быть *сжатым*. Это означит, что такое представление должно быть *существенно* более коротким, чем исходное. Это становится особенно актуальным при обработке значительных объёмов экспериментальных данных: короткие временные ряды или их сжатые структурные описания проще обрабатывать и сравнивать между собою. В то же время, такое представление должно содержать только наиболее важную для анализа информацию. Это позволяет получать наглядное (визуальное) представление имеющихся экспериментальных данных. Таким образом, можно определить (и наглядно представить) свойства каждого анализируемого временного ряда, произвести классификацию анализируемых временных рядов и, затем, выбрать те методы, которые необходимо использовать для анализа временных рядов. В соответствии с тем, мера сходства/различия временных рядов и их представлений, а также выбранный метод классификации (кластеризации) временных рядов, задают определённый вид препроцессинга.

Во-вторых, такое представление должно быть инвариантным относительно небольших вариаций исходных данных (или должно приводить к инвариантному представлению анализируемых временных рядов). Это значит, что препроцессинг временных рядов должен использоваться, прежде всего, для того, чтобы исключить влияние на результаты математического анализа временных рядов случайных факторов и привходящих обстоятельств. Инвариантность предполагает задание определённого класса преобразований исходных данных (временных рядов), относительно которого некоторое представление анализируемых временных рядов является инвариантным. Таким образом, задавая тот или иной класс преобразований исходных данных, который связан с различными свойствами анализируемых временных рядов, можно строить различные инвариантные представления, описывающие только те особенности временных рядов, которые имеют значение для решения конкретной задачи. В соответствии с этим, класс преобразований и вид инвариантного представления (а это может быть отдельная числовая характеристика, числовой вектор, числовая матрица, а может быть и структурное описание в том или ином виде) задаёт определённый вид препроцессинга.

В-третьих, такое представление должно быть пригодным для эффективного применения методов машинного обучения. Это значит, что конечное представление анализируемых временных рядов должно содержать набор потенциально информативных признаков и/или структурных описаний, которые могут быть использованы в качестве входных данных для специализированных алгоритмов машинного обучения. Таким образом, основное назначение методов препроцессинга временных рядов состоит в подготовке данных для последующего применения методов машинного обучения, и их основная специфика заключается именно в том, какова природа обрабатываемых данных. В соответствии с этим, выбор подходящего метода машинного обучения (а это могут быть и нейронные сети, и деревья решений, и самоорганизующиеся карты, и системы опорных векторов, и системы байесовского вывода) и соответствующей ему формы представления исходных данных задают определённый вид препроцессигна.

Рассмотрим подробнее формальную строну вопроса.

Предположим, имеется временной ряд, содержащий N отсчётов:

$$X = \left(x_i\right)_{i=1}^{N}.$$

Как правило, этот временной ряд является результатом преобразования непрерывного (аналогового) сигнала: дискретизация непрерывного сигнала по времени и квантование полученного дискретного сигнала по уровню приводят к формированию *цифрового* сигнала. В этом смысле, временной ряд X — это, лишь, один из возможных вариантов реализации цифрового представления непрерывного сигнала, и, вообще говоря, могут быть получены и другие представления того же непрерывного сигнала. В то же время, при анализе временных рядов, целесообраз-

но иметь дело не с конкретными реализациями, а с *инвариантными представлениями* (анализируемых временных рядов), которые не содержат индивидуальных особенностей каждой конкретной реализации. Кроме того, непрерывный сигнал может быть (если он не является реализацией детерминированного процесса) одной из возможных реализаций случайного процесса. Таким образом, отдельного рассмотрения требуют вопросы *оцифровки* временных рядов, связанные с заданием тех или иных равномерных или неравномерных сеток значений как по времени, так и по уровню.

Рассмотрим некоторое число $K < N$ и ряд, содержащий K отсчётов:

$$Y = \left(y_j \right)_{j=1}^{K}.$$

Будем называть ряд Y *производным* по отношению к исходному временному ряду X. В обычной ситуации, производный ряд Y интерпретируется как временной. В самом простом случае, такая ситуация возникает при сегментации временных рядов, когда y_j — это набор числовых характеристик, вычисляемых на сегменте с номером j ($j = \overline{1, K}$), где K — это количество сегментов, на которые разбивается временной ряд. В более сложном случае, различные элементы производного временного ряда можно рассматривать как структурные элементы анализируемых временных рядов. Это могут быть типовые фрагменты, из которых состоит исходный временной ряд, и, тогда, производный временной ряд описывает разложение исходного временного ряда на (возможно, перекрывающиеся) фрагменты. Если, однако, производный ряд не интерпретируется как временной, то он может рассматриваться как *спектр*, представляющий собою результат разложения исходного временного ряда на *составляющие*. В этом случае, каждый элемент производного ряда описывает вклад соответствующей составляющей. Гистограмма, которая описывает распределение выборочных значений, также может интерпретироваться как спектр (особенно, если рассматривать гистограммы распределений типовых фрагментов различных порядков). Чтобы получить наиболее общее представление анализируемого временного ряда, необходимо использовать матричное представление вида

$$Y = \left(y_j^{[k]} \right)_{j=1,k=1}^{K,M},$$

где каждый столбец соответствует определённому структурному элементу (типовому фрагменту), а каждая строка — определённому моменту времени (или временному интервалу).

В самом простом виде, *препроцессинг временных рядов* — это преобразование, которое позволяет перейти от заданного временного ряда X, к производному ряду Y, который в том или ином смысле *лучше*, чем исходный. Критерий оптимальности, который используется при построении производного ряда, по своей сути и определяет вид препроцессинга. Этот критерий фактически формализует наши представления о том, что является существенным для анализа имеющихся временных рядов, а что следует считать несущественным и необходимо исключить из анализа.

В том случае, когда требуется получить значение отдельной числовой характеристики ($K = 1$) или целый набор числовых характеристик (когда K относительно мало), которые описывают временной ряд в целом, рассматриваются статистические характеристики, и, тогда, число K определяет количество статистических моментов, которые могут быть использованы для представления анализируемых временных рядов. Случай $K = 1$ особый. Здесь, обычно, целесообразно строить определённую шкалу, которая позволяет классифицировать временные ряды, обладающие различными свойствами. Такая шкала называется *идентификационной*. Существует пример *фрактальной идентификационной шкалы*, которая используется для классификации временных рядов, являющихся реализациями случайных процессов с различными распределениями [8].

Если $K = N$, то наилучшим «приближением» будет сам временной ряд X. В этом случае, никакое преобразование не требуется. В то же время, преобразование, осуществляющее низкочастотную фильтрацию, позволяет перейти к временному ряду, содержащему такое же количе-

ство отсчётов, что и исходный временной ряд, и, если высокочастотные составляющие не относятся к полезному сигналу, то новое представление будет лучше, чем исходное. Если же речь идёт о простом сглаживании с использовании метода скользящего среднего, то оно осуществляется при помощи скользящего окна некоторой фиксированной ширины в m отсчётов. В этом случае, производный временной ряд будет содержать $K = N - m + 1$ отсчётов. Чтобы производный временной ряд содержал такое же количество отсчётов, что и исходный, временной ряд X необходимо дополнить отсчётами в самом начале временного рада (перед первым отсчётом временного ряда X) и — в самом конце временного ряда (после последнего отсчёта временного ряда X). Дополнение временного ряда X новыми отсчётами позволяет непосредственно сопоставлять друг другу отсчёты исходного временного ряда X и соответствующего ему производного временного ряда Y.

Строго говоря, возможен и такой вариант реализации препроцессинга, при котором количество отсчётов производного временного ряда оказывается большим, а, иногда, и существенно большим, чем количество отсчётов исходного временного ряда. На самом деле, такой вид препроцессинга может иметь смысл, поскольку любой произвольно взятый временной ряд является конечной выборкой данных из непрерывного сигнала. Дополняя временной ряд новыми отсчётами (вставляя новые отсчёты между изначально заданными) в соответствии с некоторым правилом, мы, фактически, делаем некоторые предположения относительно природы или структуры анализируемого временного ряда. Для этого, обычно, используются модели подходящего вида, которые позволяют заполнить промежутки между заданными отсчётами. Примером такого подхода является теория отсчётов сигналов, основанная на применении базисных сплайнов [9].

Совершенно другой вид препроцессинга возникает в том случае, если рассматриваются те или иные системы *характерных точек*. В качестве характерных точек могут выступать точки экстремумов, точки перегиба, точки т.н. «разладок», а также *опорные точки* — точки, структура которых является инвариантной относительно некоторого заданного класса преобразований. В этом случае, в качестве производного временного рада может выступать: сама последовательность характерных точек; последовательность интервалов между соседними характерными точками или *приращений*; последовательность структурных представлений, описывающих некоторые подпоследовательности характерных точек. Тривиальный пример производного временного ряда подобного вида — это т.н. «дискретная производная» временного ряда, определяемая как последовательность приращений или, что то же самое, разностей соседних значений исходного временного ряда. Менее тривиальный пример — это последовательность, состоящая из нулей и единиц и определяемая некоторым заранее заданным пороговым значением, в соответствии с которым, каждый отсчёт исходного временного ряда представляется единицей, если текущий отсчёт превышает заданный порог, в противном случае, текущий отсчёт представляется нулём. Такое представление кажется искусственным, однако, существуют процессы, для которых такой «искусственный» подход оказывается оправданным. Более того, существует целая математическая теория «выбросов» случайных процессов, и в рамках этой теории получены важные результаты [10].

Сегментация временных рядов

При проведении препроцессинга особую роль играет предварительная сегментация временных рядов. *Сегментация* — это разбиение временного ряда на отрезки или *сегменты*. Регулярная сегментация — это разбиение временного ряда на сегменты одной и той же заранее выбранной длины. Нерегулярная сегментация — это разбиение временного ряда на сегменты, каждый из которых имеет свою собственную длину. Особый случай — это *адаптивная* сегментация, длина каждого сегмента которой зависит от локальных свойств анализируемого временного ряда. Если допустить возможность перекрытия различных фрагментов, то такую сегментацию можно называть *фрагментацией*, подразумевая, что речь, в общем случае, идёт не о разбиении временного ряда, а о построении системы фрагментов, так или иначе соответствующих

анализируемому временному ряду. При таком обобщённом подходе необходимо в явной форме описывать алгоритм извлечения отсчётов, используемый при формировании фрагментов.

Предположим, что исходный временной ряд X, содержащий n отсчётов, разбит на K фрагментов, каждый из которых содержит N отсчётов. В этом случае, $n = NK$. Произведённая сегментация автоматически порождает матрицу

$$X_{N,K} = \left[x_i^{(j)} \right]_{i=1,j=1}^{N,K}.$$

Столбцы матрицы $X_{N,K}$ соответствуют различным сегментам заданной сегментации. При этом, сам временной ряд X можно рассматривать либо как один «длинный» сегмент (в этом случае $n = N$ и $K = 1$), либо как набор сегментов длины 1 (в этом случае $N = 1$ и $K = n$). В более общем случае, имеет смысл рассматривать фрагменты или отрезки временных рядов, каждый из которых содержит N отсчётов; при этом, сами фрагменты располагаются с некоторым шагом или сдвигом в m отсчётов друг относительно друга. Если величина сдвига меньше выбранной длины фрагментов ($m < N$), то различные фрагменты накладываются друг на друга. Если величина сдвига равна выбранной длине фрагментов ($m = N$), то речь идёт об обыкновенной сегментации временного ряда X. Если, однако, величина сдвига превышает выбранную длину фрагментов ($m > N$), то между различными фрагментами возникают «пустоты» или «зазоры», которые, фактически, исключаются из анализа. В самом общем случае, отдельные отсчёты, образующие каждый фрагмент, не обязаны непосредственно следовать друг за другом в исходном представлении временного ряда, а могут браться с некоторым шагом или сдвигом друг относительно друга.

Предположим, также, что на каждом сегменте вычисляется один и тот же набор числовых характеристик, и количество различных характеристик равно L. Произведённая сегментация автоматически порождает матрицу

$$Y_{K,L} = \left[y_i^{(j)} \right]_{i=1,j=1}^{K,L}.$$

Строки матрицы $Y_{K,L}$ соответствуют различным сегментам, а столбцы — различным числовым характеристикам. Если $L = 1$, то представление сводится к вектор-столбцу, который описывает изменение одной и той же характеристики от сегмента к сегменту. Если $L > 1$, то представление оказывается существенно многомерным. В частности, такая ситуация может возникнуть, если изначально рассматривается обработка многомерных временных рядов, каждое измерение которых может соответствовать одному из нескольких одновременно регистрируемых физиологических сигналов. Особая ситуация возникает в том случае, когда рассматриваются различные варианты вычисления одного и того же признака (при различных значениях управляющих параметров используемого алгоритма).

Выбор длины сегмента является определяющим для ряда методов математического анализа временных рядов. Если сегментов достаточно много, то статистические характеристики такого производного временного ряда можно рассматривать в качестве интегральных показателей, описывающих исходных временной ряд в целом. В то же время, слишком большое количество сегментов приводит к тому, что сами сегменты оказываются короткими. С одной стороны, в ряде случаев, необходимо, чтобы сегменты были достаточно короткими, например, если требуется, чтобы каждый сегмент можно было рассматривать как реализацию стационарного процесса. С другой стороны, сегменты должны быть достаточно длинными, чтобы характеристики, вычисляемые на каждом сегменте, имели практический смысл. Кроме того, если необходимо исследовать именно динамику изменения какой-либо числовой характеристики, то количество сегментов не должно быть слишком большим, поскольку, в этом случае, сами сегменты будут слишком короткими, и тогда реальное поведение наблюдаемой характеристики будет затушёвано случайными локальными флуктуациями. В то же время, количество сегментов не может быть и слишком маленьким, поскольку, в этом случае, длина сегмента будет сравнима с длиною целого временного ряда, и сегментация потеряет свой смысл. Если, однако, допустить частичное перекрытие сегментов, то можно увеличить общее количество сегментов, сохранив

выбранную длину самих сегментов. При этом, основным управляющим параметром оказывается величина сдвига сегментов друг относительно друга.

Сегментация временных рядов, в итоге, приводит к тому, что анализируемый временной ряд оказывается определённым образом *размеченным*. Это означает, что каждый сегмент заданной сегментации помечается именем того класса, к которому этот сегмент отнесён в результате некоторой классификации. Выбранные сегментация (регулярная или нерегулярная, с перекрытием или без перекрытия, адаптивная/неадаптивная) и способ классификации определяют способ разметки временных рядов. Если используются различные способы разметки, то временной ряд можно представлять в виде многомерного временного ряда, состоящего из имён классов, фигурирующих в различных классификациях.

Если каждому классу выбранной классификации сопоставить букву некоторого алфавита, то, в результате, каждый временной ряд можно представить в виде последовательности букв алфавита, который либо задаётся заранее, либо автоматически формируется в ходе классификации сегментов. Символический анализ временных рядов [11] непосредственно опирается на методы нелинейной динамики [12] и, в частности, на теорему Такенса о вложении [13]. Символический анализ предполагает разбиение фазового пространства на фрагменты, каждому из которых сопоставляется определённая буква. При правильном разбиении появляется возможность построить математическую модель, которая стоит за временным рядом. Эта модель, по своей сути, является марковской моделью со скрытыми состояниями [14].

Ещё один подход к классификации фрагментов временных рядов заключается в выделении т.н. «мотивов» — фрагментов, появление которых в анализируемом временном ряде является значимым [15]. Этот подход основан на двойной сегментации: сначала задаётся некоторое (например, нормальное) распределение значений и задаётся разбиение диапазона изменения данной величины на отрезки одинаковой вероятности. Каждому элементу разбиения назначается своя буква и, тем самым, формируется определённый алфавит. Затем, производится уже разбиение на сегменты самого временного ряда: на каждом сегменте вычисляется среднее значение, и уже среднее значение оцифровывается при помощи заранее заданного алфавита. Наконец, анализируя полученные в результате двойной сегментации символические последовательности, можно выделить повторяющиеся фрагменты. «Мотивы» оказываются устойчивыми конфигурациями точек, что позволяет использовать основанный на них метод классификации фрагментов временных рядов вместо элементарного алгоритма классификации фрагментов (основанного на простом скользящем окне), результаты которого являются, по своей сути, случайными или бессмысленными [16].

Совершенно другой подход к анализу временных рядов основан на понятии сложности [17]. Этот подход заключается в разбиении анализируемых временных рядов на «простые» и «сложные» фрагменты: «простые» фрагменты — это фрагменты, которые соответствуют стационарным процессам; «сложные» фрагменты — это фрагменты, соответствующие переходным процессам. Также возможна и такая интерпретация фрагментов: «простые» фрагменты представляют собою некий «фон», соответствующий определённому *типовому* поведению, а «сложные» фрагменты — это какие-либо локальные возмущения или артефакты, которые значимым образом выделяются на общем «фоне». Для описания «простых» фрагментов могут быть использованы (относительно) простые модели, в то же время «сложные» фрагменты определяются как такие множества, при исключении которых качество аппроксимации «простых» фрагментов оказывается наилучшим. Несколько иной, алгоритмический подход к понятию сложности связан с теорией приближений и восходит к идее А.Н.Колмогорова описывать «сложность» объекта длиной программы, которая необходима для его построения [18].

Главные компоненты временных рядов

Сегментация временных рядов тесно связана с преобразованием одномерных временных рядов в многомерный временной ряд по методу «Гусеница», а именно: задаётся величина N, которая называется *длиной* «*гусеницы*», и формируется соответствующая заданной длине матрица, транспонированная к матрице $X_{N,K}$. Строки полученной матрицы описывают различные

положения «гусеницы», последовательно перемещающейся вдоль отсчётов временного ряда. Всего имеется $K = n - N + 1$ строка и N столбцов. Это позволяет интерпретировать данную матрицу как многомерный временной ряд размерности N. В частности, эту матрицу можно использовать для построения скользящего среднего исходного временного ряда с окном, ширина которого совпадает с длиной «гусеницы». Дальнейшая обработка предполагает сингулярное разложение данной матрицы и построение главных компонентов анализируемых временных рядов. Отбирая главные компоненты и осуществляя обратное преобразование, можно восстановить исходный временной ряд (если используются все главные компоненты) или получить некоторое представление, которое является лишь приближением к исходному ряду, но обладает некоторыми хорошими свойствами. Таким образом, сравнение исходных временных рядов может быть сведено к сравнительному анализу соответствующих им главных компонентов. Длина «гусеницы», задающая размерность многомерного представления временного ряда, определяет свойства всех матриц и количество различных главных компонентов. Это значит, что при проведении сравнительного анализа имеет смысл изначально фиксировать длину «гусеницы». В то же время, сравнительный анализ результатов, полученных для различных длин «гусеницы», требует особой осторожности.

Если, теперь, рассмотреть некоторую регулярную сегментацию с длиной сегментов, составляющей M отсчётов, то можно обнаружить, что, каждому сегменту исходного временного ряда будет соответствовать сегмент многомерного временного ряда или соответствующей ему матрицы. Этот сегмент, очевидно, будет содержать $m = N - M + 1$ отсчёт ($m = N - M + 1$ строку). При этом, в зависимости от того, какая схема сегментации используется (без перекрытия или с перекрытием), будет возникать аналогичная сегментация соответствующего многомерного временного ряда (соответствующей матрицы). Рассматривая изменение главных компонентов от сегмента к сегменту, можно получить представление о динамических свойствах анализируемого временного ряда. В то же время, фиксируя длину «гусеницы» и рассматривая различные длины сегментов, взятые из некоторого заранее заданного набора допустимых длин, можно получить различные наборы главных компонентов и соответствующих им собственных чисел. Применяя, затем, к полученным данным методы машинного обучения, можно выбрать ту длину сегментов, которая является наилучшей.

Выбор главных компонентов для представления анализируемого временного ряда связано с оценкой величин собственных чисел, соответствующих главным компонентам. Вычисление собственных чисел для всех сегментов анализируемого временного ряда приводит к формированию матрицы $Y_{K,L}$, число столбцов которых совпадает с заданной длиной «гусеницы». Собственные числа, имеющие малую величину, являются *несущественными*, а соответствующие им главные компоненты исключаются из анализа. Задавая равномерную или неравномерную сетку пороговых значений, а также требуемое количество *существенных* компонентов, можно определить те сегменты, которые обеспечивают заданное количество существенных компонентов для каждого порогового значения. Если, теперь, рассматривать эту многомерную зависимость для объектов определённого класса (например, для определённого класса испытуемых), то можно определить долю «правильных» сегментов, удовлетворяющих заданным требованиям.

Таким образом, сегментация, особенно в том случае, когда рассматривается не какая-то одна произвольно взятая длина сегмента, а используется несколько различных длин сегментов из определённого набора допустимых (и тем или иным образом обоснованных) значений, становится важным инструментом анализа временных рядов.

Заключение

Цель обработки временных рядов — та или иная классификация анализируемых временных рядов. В наилучшей ситуации, классификация временных рядов полностью соответствует тем функциональным состояниям, в которых могут находиться испытуемые. В этом случае, принцип, положенный в основание классификации (он же критерий классификации), оказывается основанием для построения (относительно) простого решающего правила. Для построения про-

стых решающих правил используются числовые характеристики, которые связаны с теми или иными аспектами анализируемых временных рядов. Если, например, в качестве основной числовой характеристики рассматривается фрактальная размерность, то может быть, в частности, проверена гипотеза о том, что ФС «в норме» соответствуют процессы с более простым управлением, а, значит, и меньшей фрактальной размерностью, в то время как ФС «в патологии» соответствуют процессы с более сложным управлением, а, значит, и большей фрактальной размерностью. Это означает, что даже вычисление единственного информативного признака может быть основанием для построения хорошего решающего правила.

Более распространённой является ситуация, когда такого информативного признака нет или, когда, не смотря на наличие одного или нескольких информативных признаков, имеющихся априорных сведений недостаточно для точной и надёжной классификации анализируемых временных рядов. В этой ситуации необходимо иметь некоторую меру сходства временных рядов (или аналогичную ей меру различия) и использовать именно эту меру для классификации, а, точнее, кластеризации временных рядов. Для каждой меры сходства и для каждого алгоритма кластеризации можно получить свой набор кластеров. Рассматривая различные меры сходства и различные алгоритмы кластеризации можно получить развёрнутое представление экспериментальных данных в виде размеченной матрицы экспериментальных данных. Дальнейшее применение методов машинного обучения к довольно подробно размеченным экспериментальным данным способно дать более глубокие и надёжные результаты, чем непосредственное применение методов машинного обучения «в лоб» к данным, которые предоставляются для обработки «как есть» в своём исходном виде.

Методы препроцессинга используются именно для того, чтобы предоставить данные для машинного обучения в виде, наиболее удобном для применения алгоритмов обработки экспериментальных данных. Рассмотрение различных подходов к анализу временных рядов и реализующих их алгоритмов позволяет выделить определённые элементы, являющиеся общими для различных методов анализа, использовать эти общие элементы для реализации препроцессинга ТТГ.

Важную роль при реализации препроцессинга анализируемых временных рядов играет сегментация. Во-первых, сегментация является неотъемлемой частью ряда алгоритмов обработки временных рядов. Так что, её применение является, что называется, *стандартным* этапом математического анализа временных рядов, когда необходимо получить более подробное и, вместе с тем, более сжатое представление об анализируемом временном ряде. Во-вторых, сегментация позволяет существенным образом увеличить количество объектов, участвующих в распознавании, и количество потенциально информативных признаков, участвующих в построении решающих правил. При этом, задание системы пороговых значений и дополнительного (внешнего) критерия отбора «правильных» сегментов, позволяет строить гибкие решающие правила, допускающие управление уровнем обобщающей способности. В-третьих, рассмотрение сразу нескольких различных сегментаций, производимых для различных длин сегментов из заранее заданного набора, позволяет получать зависимости потенциально информативных признаков от длины сегмента. (Например, вычисление показателя Хёрста [19], используемого для оценки фрактальной размерности, осуществляемое по оригинальному алгоритму [7], требует предварительной сегментации временного ряда и построения зависимости основного вычисляемого отношения от длины сегмента.) В итоге, методы препроцессинга временных рядов должны использоваться для полной реконструкции структуры анализируемых временных рядов, и здесь могут применяться специализированные методы распознавания образов на временной оси [20].

ЛИТЕРАТУРА

1. Романов С.П., Якимовский А.Ф., Пчелин М.Г. Метод тензометрии для количественной оценки тремора // Физиологический журнал им. И.М. Сеченова. 1996, т. 82, № 2. С. 118–123.
2. Романов С.П., Пчелин М.Г., Якимовский А.Ф. Характеристики изометрически регистрируемого тремора при поражении экстрапирамидной системы // Российский физиологический журнал им. И.М. Сеченова. 1997, т. 83, № 3. С. 133–139.
3. Бернштейн Н.А. О построении движений. М.: Медгиз, 1947. 254 с.
4. Павлов А.Н., Анищенко В.С. Мультифрактальный анализ сложных сигналов // Успехи физических наук. 2007. Т. 177. С. 859.

5. Дик О.Е., Романов С.П., Ноздрачев А.Д. Энергетические и фрактальные характеристики физиологического и патологического тремора руки человека // Физиология человека, 2010, том 36, № 2, с. 92–100.
6. Дик О.Е., Ноздрачев А.Д. Нелинейная динамика непроизвольных колебаний руки человека при двигательной патологии // Физиология человека, 2015, том 41, № 2, с. 53–59.
7. Любушкин А.А. Фрактальный анализ временных рядов. М.: 2006.
8. Кобенко В. Ю. Фрактальная идентификационная шкала // Омский научный вестник. №3 (83), 2009. С. 205–213.
9. Свиньин С.Ф. Базисные сплайны в теории отсчётов сигналов. СПб.: Наука, 2003.
10. Тихонов В.И., Хименко В.И. Выбросы траекторий случайных процессов. М: Наука, 1987.
11. Daw C.S., Finney C.E.A., Tracy E.R.. A review of symbolic analysis of experimental data // Rev.Sci. Instrum. 2003. V. 74. PP. 915–930.
12. Кузнецов С.П. Динамический хаос (курс лекций). М.: Издательство физико-математической литературы, 2006. 356 с.
13. F. Takens. Detecting strange attractors in turbulence. In D. A. Rand and L.-S. Young. Dynamical Systems and Turbulence, Lecture Notes in Mathematics, vol. 898. Springer-Verlag. 1981, pp. 366-381.
14. Baum, Leonard E.; Petrie, Ted. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. Ann. Math. Statist. 37 (1966), no. 6, pp. 1554–1563.
15. Lin, J., Keogh, E., Lonardi, S. & Patel, P. (2002). Finding Motifs in Time Series. In proceedings of the 2nd Workshop on Temporal Data Mining, at the 8th ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining. Edmonton, Alberta, Canada, July 23-26. pp. 53-68
16. J. Lin, E. Keogh, W. Truppel, Clustering of streaming time series is meaningless, in: Proceedings f th ACM SIGMOD Workshop on Research Issues Data Mining and Knowlegde Discovery DMKD 03, 2003, p. 56.
17. Браверман Э.М. Мучник И.Б. Структурные методы обработки эмпирических данных. М.: Наука, 1983. 464 с.
18. Дарховский Б.С., Каплан А.Я., Шишкин С.Л. О подходе к оценке сложности кривых (на примере электроэнцефалограммы человека) // Автоматика и телемеханика. 2002. № 3. С. 134–140.
19. Ю. А. Калуш, В. М. Логинов Показатель Хёрста и его скрытые свойства // Сибирский журнал индустриальной математики. 2002.Т. 5. Вып. 4. С. 29–37.
20. Моттль В.В., Мучник И.Б. Детерминистские модели и методы распознавания образов на оси времени. I. Основные модели и разрешимость задачи распознавания // Автоматика и телемеханика. 1991. № 3. С. 120–132.

O.V. Zhvalevsky (Federal State Institution of Science St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg)

The applying of time series preprocessing for tenzotremorogramms processing

The paper is concern with problem of tenzotremorogramms preprocessing that appears in Parkinson disease diagnostics. Time series preprocessing is followed by machine learning. The main task of tenzotremorogramms preprocessing is proposed. The main species of time series preprocessing are described. Also time series segmentation using for time series preprocessing representation is proposed.

ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ И СИСТЕМА МОНИТОРИНГА КОНКУРЕНТНОЙ СПОСОБНОСТИ ПРОДУКЦИИ КАК ГЛАВНОГО ФАКТОРА ТЕХНОЛОГИЧЕСКОГО РАЗВИТИЯ ПРЕДПРИЯТИЯ

А.В. Алексеев, В.Ю. Бобрович, В.В. Антипов, А.В. Смольников
(*Институт автоматизации процессов борьбы за живучесть корабля, судна, С.-Петербург, 198262, Ленинский пр., 101, оф. 162, iapbgks@bk.ru*)

Аннотация

Рассмотрены информационная технология и система мониторинга конкурентной способности продукции и/или услуг как главный фактор технологического развития предприятия. Методической основой анализа и мониторинга признано считать системно значимые оценки полимодельного агрегированного показателя качества, конкурентной способности, перспективности развития продукции и/или услуг, а также предприятия в целом.

Предложение иллюстрировано примером реализации с использованием программного комплекса «КСПР-Ф» по оценке и мониторингу конкурентной способности и перспективности развития пяти объектов морской техники класса «Большой противолодочный корабль» (при глубине анализа 52 года), составивших порядка 42%, что признано весьма существенным для кораблей данного класса.

Отмечена исключительная полезность подобного анализа и синтеза объектов и систем для выявления направлений перспективного технологического развития современных сложных объектов и систем, для обоснования и принятия соответствующих проектных и управленческих решений, прежде всего, в рамках концептуального и исследовательского системно-технического проектирования, системного анализа тенденций и оптимизации программ развития.

Ключевые слова: критический объект, квалиметрическое обеспечение, системный мониторинг, интеллектуальная поддержка, агрегированный показатель качества.

Введение

Наибольшее воздействие на конкурентную способность выпускаемой продукции, услуг и предприятия в целом, возможность вести успешное конкурентное противоборство на рынке, могут и оказывают инновационные, инвестиционные и финансовые факторы деятельности предприятия [1, 2].

При этом, основными системными факторами конкурентной способности товара и/или услуг, а также перспективности развития предприятия, как правило, являются:

- качество товаров и услуг, определяемое действующими стандартами, нормами и рекомендациями заказчиков и потребителей, а также фактическими условиями конкуренции;
- экономические показатели, формирующие себестоимость и цену товаров и/или услуг;
- организационно-экономические показатели соответствия производственных программ производственному потенциалу;
- конкурентная способность самого предприятия, его финансово-экономического состояния и репутации на рынке.

Вместе с тем, анализ практик оценки эффективности инвестиционных проектов и их привлекательности по критериям PV, PI, NPV, PBP, IRR, MIRR, PP, DPP, ARR показывает, что показатели конкурентной способности при этом практически не учитываются [3, 4]. Это существенно ограничивает возможности инвестиционного анализа, синтеза и оптимизации проектных и управленческих решений, технологического развития предприятий.

Предлагаемые результаты исследований посвящены учету факторов и показателей конкурентной способности при оценке эффективности инвестиционных проектов. Это позволит переходить конкретно от результатов анализа к синтезу путей повышения эффективности инвестирования и инновационной привлекательности технических, организационных предложений, мотивированию сотрудников на поиск оптимальных направлений и вариантов совершенствования выпускаемой продукции, а также перспективных путей развития предприятий и организаций в целом.

Конкурентная способность, как главный фактор технологического развития предприятия

Постановка задачи. Опыт системного анализа сложных организационно-технических производственных процессов, их мониторинга и управления, а также инновационных и инвестиционных проектов показывает, что в настоящее время в условиях резкого возрастания функциональной, структурной, организационной и организационно-технической сложности выпускаемой продукции и/или услуг все большее значение приобретают системные показатели их качества [5, 6, 7]. Более того, с использованием возможностей современных информационных технологий и систем - к их непрерывному наблюдению (мониторингу) с переходом к количественной оценке и мониторингу их конкурентной способности (КС), а для предприятия (организации, компании, корпорации) в целом - к перспективности развития (ПР).

В развитие новых метрик КС и ПР, преимущественно, для риск-ориентированных технологий обеспечения безопасности ниже рассмотрено направление, в основу которого положено оценивание и мониторинг, прежде всего, системных показателей качества.

В этой связи признано целесообразным обосновать выбор наиболее предпочтительного критерия и модели оценки показателя с названными системными свойствами.

Предлагаемое решение задачи. С этой целью на основе методологии, модели и алгоритма оценки полимодельного агрегированного показателя качества (АПК) сложных организационно-технических объектов [6, 7, 8] предложено конкурентную способность продукции и/или услуг оценивать с использованием гармонической модели учета конкурентного превосходства продукции по частным показателям качества (ЧПК) в сравнении с аналогичными для базового объекта, агрегируемых в групповые показатели качества (ГПК) и далее - в модельный показатель (МПК) и, наконец, в АПК с учетом индексов критериальной значимости (ИКЗ) и индексов характера влияния (ИХВ) ЧПК, ГПК и МПК на полимодельный АПК.

Методика решения задачи. Методически оценка комплекса названных показателей с их сведением к единому системному и полимодельному АПК производится на основе соответствующих программных комплексов вычислительной поддержки операторов с формированием системы взаимосвязанных показателей и их предпочтений.

Это позволяет, как подтверждено многочисленными вариантами числового моделирования, в том числе при альтернативном сравнении семи вариантов систем интеллектуальной поддержки принятия управленческих решений капитана транспортного судна, от задаваемых количественных значений ЧПК перейти к оценке КС продукта и/или услуги [6, 7, 8]. И, аналогично, с учетом значимости каждого из их видов - ко всему комплексу выпускаемой продукции и/или услуг предприятия.

Полимодельная специфика предложения. В сочетании с финансово-экономическими показателями, оцениваемыми, например, с использованием программных комплексов типа «Альт Инвест», предложенный метод позволяет анализировать взаимосвязь эффективности и привлекательности инновационных проектов с их конкурентной способностью как главным фактором управления технологическим развитием предприятия, а в перспективе – ведомства, отрасли, государства.

Возможности предложения. Рассматриваемый метод мониторинга конкурентной способности выпускаемой продукции и/или услуг позволяет количественно обосновывать, ранжировать и оптимизировать управленческие инвестиционные решения, обеспечивать «прозрачное» прогнозирование инновационного и инвестиционного развития предприятия, управления его технологическим развитием.

Пример реализации предложения. С использованием программного комплекса «КСПР-Ф» и доступных в открытой литературе данных на рисунке приведены результаты сравнительной оценки и анализа КС и ПР объектов морской техники (ОМТ) класса «Корабль: БПК» (большой противолодочный корабль). Квалиметрическое ранжирование пяти вариантов ОМТ показало эволюционное наращивание качества и перспективность развития за период с 1966 г. по 2018 г. (при глубине анализа 52 года) порядка 42%, что весьма существенно для кораблей данного класса [3, 6].

Оценка перспективности развития ОМТ класса "Корабль: БПК"

Назначение ОМТ:	*Обеспечение боевого превосходства в составе ТГ, ОТС.*	
Предназначение ОМТ класса "Корабль: БПК"	*Цель: Борьба с подводными лодками (ПЛ) в Океанской зоне*	
Основные свойства ОМТ класса "Корабль: БПК" варианта "**5.Ф-22350 "АФСС Горшков"** по отношениию к варианту **"1.БПК-61"Строгий"**:	**1. Боевой потенциал (качество достижения цели, ПС "1.Экипаж" (БГ, ЧФ, компетенции, освоение), "2.Управление", "3.Навигация, связь, сигнализация", "4.Целевая"):** 1.1.Управляемость. 1.2.Вооруженность. 1.3.Тактическое преимущество.	**42,4%**
	2. Боевая устойчивость (ПС "8.Маневрирование", "9.Стабилизация"): 2.1.Скрытность. 2.2.Живучесть. 2.3.Восстанавливаемость (экипажа, боевая, информационная, техническая).	**87,4%**
	3. Технологичность применения (ПС "5.Корпус и надстройки", "6.Движение", "7.Энергия"): 3.1.Мореходные свойства (непотопляемость, остойчивость, плавучесть, прочность, ходкость, управляемость). 3.2.Энерговооруженность. 3.3.Безопасность службы, эксплуатации.	**24,7%**
	4. Ресурсная обеспеченность: 4.1.Оперативная обеспеченность. 4.2.Экономичность владения (Стоимость закупки, освоения,эксплуатации, расходных материалов, ремонта). 4.3.Пополняемость запасами.	**-16,1%**
	5. Перспективность и др.: 5.1.Боевое превосходство. 5.2.Перспективность развития. 5.3.Другие.	**30,4%**



Боевое превосходство (конкурентная способность) по отношению к варианту 1, БП (КС), %:	**42,4%**
Перспективность развития по отношению к варианту 2 (ПР), %:	**42,0%**

Критерии оценки качества \ Варианты ОМТ	ИКЗ (вес) %	Z	1.БПК-61 "Строгий" (Базовый	2.БПК-1134-А "Кронштадт"	3.БПК-1134-Б "Керчь"	4.БПК-1155 "Адм.Ч"	5.Ф-22350 "Горшков"
Год введения в строй			5.5.66	12.12.66	5.1.76	28.1.99	?.?.2018
БОЕВОЙ ПОТЕНЦИАЛ, БПапк, ед.	-	-	**1,00**	**1,00**	**1,04**	**1,18**	**1,43**
БС по мультипликативной модели, БПм, ед.	-	-	**1,00**	**0,96**	**0,98**	**1,09**	**1,31**
БС по аддитивной модели, БПа, ед.	-	-	**1,00**	**1,05**	**1,10**	**1,27**	**1,55**
1. БОЕСПОСОБНОСТЬ, КП, ед.	**20%**	**1**	**1,01**	**1,77**	**1,87**	**2,39**	**3,03**
1.1.Управляемость, %	30%	1	1,00	0,97	1,00	1,13	1,32
1.2.Вооруженность, %	50%	1	1,00	2,76	2,99	4,21	5,75
1.3.Тактическое превосходство, %	20%	1	1,06	1,05	1,06	1,06	1,06
2.БОЕВАЯ УСТОЙЧИВОСТЬ, КП, ед.	**20%**	**1**	**1,00**	**1,02**	**1,20**	**1,45**	**1,87**
3.ТЕХНОЛОГИЧНОСТЬ ПРИМЕНЕНИЯ, КП, ед.	**20%**	**1**	**1,00**	**1,01**	**1,05**	**1,11**	**1,25**
4.РЕСУРСНАЯ ОБЕСПЕЧЕННОСТЬ, КП, ед.	**20%**	**1**	**1,00**	**0,80**	**0,73**	**0,72**	**0,84**
5.ПЕРСПЕКТИВНОСТЬ и др. факторы, КП, ед.	**5%**	**1**	**1,00**	**1,01**	**1,04**	**1,15**	**1,30**
9.Экономичность владения, млн.руб.	**10%**	**1**	**0,32**	**0,10**	**0,09**	**0,09**	**0,12**
9.1.Закупочная стоимость, млн.руб.	50%	-1,0	50	150	100	150	250
9.2.Расходные материалы на 1т.км, млн.руб.	40%	-1,3	10	30	20	30	40
9.3.Стоимость обслуживания, млн.руб.	**10%**	-3,0	0,7	0,9	0,7	0,8	1,2
11.Другие факторы развития, уе	**5%**	1	1,0	1,0	1,0	1,0	1,0
Боевое превосходство (конкурентная способность), БП, %			1,00	1,00	1,03	1,18	1,42
Ранжирование оценок, **R**			**5**	**4**	**3**	**2**	**1**
Перспективность развития (ПР) в сравнении с вариантом 2, %			1,00	1,00	1,03	1,17	1,42
ПК "КСПР-Ф" Эксперты:			Алехин К.В.		Карпов А.Е.		*20.09.18*

Рис. 1. Пример реализации сравнительной оценки и анализа объектов морской техники класса «Корабль: БПК»

При этом, впервые показана возможность комплексной оценки свойств ОМТ типа: 1.Боевой потенциал (на рис. 1 приведены агрегируемые при этом ГПК). 2.Боевая устойчивость. 3.Технологичность применения. 4.Ресурсность обеспечения. 5.Перспективность и др.

Следует отметить возможность в рамках иллюстрируемой методологии различных вариантов агрегирования показателей качества, что в общем случае является прегеротивой специализированных экспертных комиссий. Это позволит соответственно уточнить получаемые оценки и результаты. Но при этом остается бесспорной, по мнению авторов, исключительная полезность подобного системного анализа и синтеза свойств объектов анализа как для выявления направлений перспективного технологического развития ОМТ, так и для обоснования и принятия соответствующих проектных решений. Прежде всего, в рамках концептуального и исследовательского системно-технического проектирования, анализа тенденций развития и оптимизации программ развития.

В целом, рассмотренное предложение и целый ряд выполненных вариантов числового моделирования для различных классов ОМТ [4, 5, 6, 9] подтвердил целесообразность и технологическую возможность оценки и мониторинга конкурентной способности продукции и/или услуг, что позволяет рассматривать представленную технологию, как главный фактор технологического развития предприятия.

Заключение

В результате выполненных исследований по анализу и синтезу информационных технологий и систем мониторинга конкурентной способности продукции как главного фактора технологического развития предприятий показана целесообразность и перспективность внедрения информационных технологий данного типа с целью исследования и результативного управления системными свойствами и характеристиками объектов.

Квалиметрическая оценка системных показателей качества (АПК, КС, ПР), их мониторинг, прогнозирование, нормирование и контроль позволяют при управлении качеством объектов обоснованно формировать и задавать направления и рубежи обеспечения как конкурентной способности и перспективности развития современных сложных объектов и систем, так и предприятий в целом, производящих соответствующую продукцию и услуги.

ЛИТЕРАТУРА

1. Горбашко Е.А. Менеджмент качества и конкурентоспособности. — СПб.: СПбГУЭФ, 2002 — стр. 104
2. Азоев Г.Л. Конкуренция: анализ, стратегия и практика — М.: Центр экономики и маркетинга, 2006. — стр. 131
3. Алексеев А.В. Квалиметрическое обеспечение организации принятия проектных и управленческих решений в сложных системах / Системный анализ при создании кораблей, комплексов вооружения и военной техники. Тематический сборник. Выпуск 15. – СПб.: «Моринтех», 2008, с. 67 - 75.
4. Алексеев А.В., Антипов В.А., Бобрович В.Ю., Смольников А.В. Оптимизация системных решений при управлении живучестью судовых автоматизированных систем и комплексов на основе анализа их конкурентной способности / Материалы 9-й конференции «Информационные технологии в управлении» (ИТУ-2016). – СПб.: АО «Концерн «ЦНИИ «Электроприбор», 4-6.10.2016 г. - Спб., ГНЦ РФ АО «Концерн «ЦНИИ «Электроприбор», 2016.
5. Алексеев А.В., Антипов В.А., Бобрович В.Ю., Евсеенко С.М. Реализация обобщенного метода квалиметрического анализа факторов развития и технология обеспечения управления развитием критических морских объектов // Морские интеллектуальные технологии. Научный журнал № 1 (31) Т.1, 2016, с. 27 – 37.
6. Бобрович В.Ю., Алексеев А.В., Антипов В.В., Смольников А.В., Кузнецов В.В., Тычинин И.Ю. Конкурентная способность как главный фактор инвестиционной привлекательности / IV Санкт-Петербургский международный экономический конгресс (СПЭК-2018): материалы. – СПб.: НИИР им. С.Ю. Витте, 2018, с. 46 – 51.
7. Алексеев А.В. Концептуальные аспекты управления развитием критических объектов морской техники и морской инфраструктуры // Морские интеллектуальные технологии. Научный журнал № 2 (28) Т.1, 2015, с. 47 – 57.
8. Алексеев А.В., Сус Г.Н., Ушакова Н.П. Системный анализ и ранжирование качества интеллектуальной поддержки принятия решений и управления борьбой за живучесть корабля, судна / Материалы 9-й конференции «Информационные технологии в управлении» (ИТУ-2016). – Спб., ГНЦ РФ АО «Концерн «ЦНИИ «Электроприбор», 2016, с. 786-790.
9. Александров В.Л., Алексеев А.В. Теория практики квалиметрического обеспечения конкурентной способности и перспективности развития объектов морской техники и морской инфраструктуры / Восьмая Всероссийская научно-практическая конференция по имитационному моделированию и его применению в науке и промышленности «Имитационное моделирование. Теория и практика» (ИММОД-2017). 18-20 октября 2017 г. Труды конференции – СПб.: НОИМ, 2017, с. 74 - 79.

A.V. Alekseev, V.Yu. Bobrowicz, V.V. Antipov, A.V. Smolnikov (Institute of automation of processes of struggle for survivability of the ship, vessel)

Information technology and monitoring system of the competitive ability of the production as the main factor of technological development of the enterprise

The information technology and system of monitoring of competitive ability of production and/or services as the main factor of technological development of the enterprise are considered. The methodical basis of the analysis and monitoring is considered to be systemically significant estimates of the polymodel aggregated indicator of quality, competitive ability, prospects of development of products and/or services, as well as the enterprise as a whole.

The proposal is illustrated by an example of implementation using the software complex "kspr-f "to assess and monitor the competitive ability and prospects of development of five objects of marine equipment of the class" Big anti-submarine ship " (at a depth of 52 years), which amounted to about 42%, which is recognized as very important for ships of this class.

The exceptional usefulness of such analysis and synthesis of objects and systems to identify areas of promising technological development of modern complex objects and systems, to substantiate and make appropriate design and management decisions, primarily within the framework of conceptual and research system-technical design, system analysis of trends and optimization of development programs.

ТРЕХЭТАПНАЯ БЛОК-СХЕМА СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ

И. Г. Ханыков[1]
(*Институт информатики и автоматизации Российской академии наук,
С.-Петербург, 199178, 14 линия, д.39, igk@iias.spb.su*)

Аннотация

Приводятся требования к актуальным алгоритмам сегментации изображений. Сопоставляются применимые в сегментации классические методы кластерного анализа: методы Уорда, Оцу, К-средних, модель Мамфорда-Шаха. Обосновывается выбор классического метода Уорда. Приводятся его достоинства и недостатки. Предлагается типовая блок-схема последовательности алгоритмов, позволяющая обойти проблему вычислительной сложности, характерную для метода Уорда. Описано применение идеи обратимых операций в обработке изображений. Описываются методы улучшения качества заданного разбиения изображения. Приводятся экспериментальные результаты по улучшению качества традиционной сегментации, по улучшению качества иерархической последовательности кусочно-постоянных разбиений и по выделению одинаковых объектов на паре различных изображениях.

Ключевые слова: кластерные методы, иерархическая сегментация изображений, метод Уорда, обратимые вычисления, улучшение качества изображения

Введение

Задача сегментации относится к стадии предварительной обработки цифровых изображений. Результаты сегментации служат исходными данными для последующего выделения объектов, распознавания признаков, анализа сцен и прогнозирования ситуации. К актуальным алгоритмам сегментации изображений (АСИ) выдвигаются следующие требования [1]:

1) Способность работать при отсутствии априорной информации об объектах интереса и о тематической сути изображения;

2) Наличие установленного критерия качества, который позволяет оценивать полученное разбиение изображения на кластеры/сегменты.

3) Возможность сегментирования/дробления изображения на любое число цветов/кластеров от 1 до N, где N – число пикселей в рассматриваемом изображении

4) Выполнение вычислений в режиме реального времени.

5) Адекватность результатов, заключающейся в соответствии выделенных сегментов/кластеров границам и областям объектов на изображении.

Из всего множества АСИ особое место занимает группа кластерных методов. Она удовлетворяет большинству приведённых требований, в частности п.1. Самыми распространенными кластерными методами, применяемыми в сегментации изображений, являются методы Уорда [2], Оцу [3], К-средних [4], модель сегментации Мамфорда-Шаха [5], [6]. В таблице 1 приводится их сопоставление.

Во-первых, ни один из методов не требует знания/наличия априорной информации об тематике изображения (п.1). Приведенные методы пригодны для обработки изображений любого содержания и тематики.

Во-вторых, приведенные методы минимизируют установленный функционал качества (п.2), значение которого служит индикатором качества разбиения изображения на кластеры пикселей, сегменты изображения. Например, метод Оцу минимизирует суммарное квадратичное отклонение; метод К-средних – суммарное расстояние до центров кластеров; модель сегментации Мамфорда-Шаха – энергетический функционал; метод Уорда – суммарное квадратичное отклонение.

Различие рассматриваемых кластерных методов проявляется в способе разбиения изображения на кластеры писклей или сегментов (п.3). Метод Оцу разбивает все пиксели только на два кластера. Метод К-средних разбивает пиксели изображения на заданное число кластеров, которое задается заранее и фиксировано на всем протяжении исполнения алгоритма. Модель сегментации Мамфорда-Шаха и метод Уорда генерируют множество разбиений исходного изоб-

[1] Младший научный сотрудник лаборатории Прикладной информатики и проблем информатизации общества

ражения от 1 до N, где N – число пикселей в исходном изображении. Но при этом, если в методе Уорда рассматривается все сочетания кластеров, то в модели Мамфорда-Шаха – только пары смежных сегментов изображения.

Таблица 1

Сопоставление классических методов кластеризации пикселей, сегментации изображений

№	Метод / Признак	Оцу	К-средних	Мамфорда-Шаха	Уорда
1	**Априорная информация**	Нет	Нет	Нет	Нет
2	**Функционал качества**	Да	Да	Да	Да
3	**Изменяемое число кластеров/ сегментов**	Нет, фиксировано, 2	Нет, фиксировано, $1<k<N$.	Да, итеративно возрастает от 1 до N.	Да, итеративно возрастает от 1 до N.
4	**Линейная вычислительная сложность**	Да, O(N)	Нет, O(n^2), где n – число кластеров	Да, O(N)	Нет, >O(N^2)
5	**Адекватность результатов**	Да, ограничено 2 кластерами.	Нет	Нет	Да

Время процесса сегментации изображения зависит от выбранного метода. Сопоставляемые методы обладают различной вычислительной сложностью и, следовательно, различной возможностью обработки изображений в режиме реального времени. Классический метод Оцу [3] разбивает пиксели за линейное время. Метод строит гистограмму распределения яркостей пикселей и за один проход разбивает все множество писклей на два кластера по условию значению минимум суммарной квадратичной ошибки. Модификации классического метода Оцу обладают гораздо большей вычислительной сложностью. Например, вычислительная сложность мультипорогового метода Оцу [7], зависящего от числа задаваемых порогов, имеет экспоненциальный характер. Время, затрачиваемое методом К-средних на разбиение изображения на фиксированное число кластеров, зависит квадратично от задаваемого значения числа кластеров. Модель Мамфорда-Шаха сегментирует изображение за линейное время близкое реальному. Вычислительная сложность классического метода Уорда тоже возрастает квадратично с ростом числа рассматриваемых кластеров писклей, что затрудняет применение метода напрямую в обработке изображений.

Выбор метода также влияет как на качество результата обработки, так и на адекватность сегментации изображения (п.5). Под адекватным результатом сегментации изображения понимаем соответствие кластеров пикселей или сегментов изображения областям и границам объектов заданного исходного изображения. Такое определение носит субъективный характер, так как понимание соответствия зависит от индивидуального восприятия пользователя. Для объективного оценивания результатов сегментации изображений в настоящей работе используется суммарное квадратичное отклонение.

В работах [8] использование суммарного квадратичного отклонения критикуется за то, что оно «дает одинаковое значение при разных типах шумов, одни из которых более, а другие менее заметны». Поэтому с точки зрения гипотезы о том, что «человеческая зрительная система очень адаптирована извлекать структурную информацию», суммарное квадратичное отклонение «не очень хорошо подходит к оценке визуально воспринимаемого качества». Данная критика уместна только в условиях задачи сопоставления различных контрольного и поверяемого изображений при заданной метрике из категории full-reference, либо reduced-reference.

Применение суммарного квадратичного отклонения оправдано по ряду причин. Во-первых, это – классическая величина и она должна рассматриваться в первую очередь. Во-вторых, оно легко в подсчете, имеет ясное физическое значение, и математически удобно в контексте оптимизации. В-третьих, суммарное квадратичное применяется в кластерном анализе для оценки качества объединения элементов в классы.

Из рассматриваемых методов к адекватным относятся метод Оцу и метод Уорда. Метод Оцу возвращает оптимальное разбиение на два кластера и применим только для полутоновых изображений, редуцируя их до бинарных. Более общим является метод Уорда. Он применим как к цветным, так к полутоновым изображениям. Метод возвращает множество разбиений, каждое последующее получается слиянием пары кластеров в один. Разбиения получаются адекватны-

ми, но общее время обработки изображения получается велико из-за полного перебора всех пар возможных сочетаний.

Метод К-средних и модель Мамфорда-Шаха относятся к категории менее адекватных. Результаты метода К-средних зависят от начального выбора центров кластеров. При их «плохом» задании метод останавливается в локальном минимуме. Зачастую результат получается неприемлемым. Модель Мамфорда-Шаха возвращает грубую сегментацию, так как на каждом шаге рассматриваются только пары смежных пикселей, сегментов.

Рассматриваемый в работе метод Уорда удовлетворяет четырём из пяти требований, приведенных к актуальным АСИ: (п.1), (п.2), (п.3), (п.5). Однако, метод Уорда обладает рядом недостатков, существенный из которых – высокая вычислительная сложность (п.4).

Настоящая работа посвящена методике нивелирования недостатков классического метода Уорда, позволяющей применять его в обработке изображений в задачах кластеризации и сегментации. Приводится идея иерархической сегментации изображений. Описывается достоинства и недостатки классического метода Уорда. Предлагается типовая блок-схема последовательности алгоритмов, обходящая недостатки классического метода посредством промежуточной обработки исходного изображения для кластеризации его пикселей. В завершении приводятся экспериментальные результаты.

Идея иерархической сегментации изображений

Идея иерархической сегментации [9] заключается в генерации последовательности кусочно-постоянных разбиений исходного изображения. Исходное изображение разбивается на вложенные изображения, каждое из которых заполнено одинаковыми пикселями с усредненным значением яркости. Так исходное изображение преобразуется в свое *кусочно-постоянное приближение*. Каждое последующее разбиение получается объединением пары кластеров пикселей, в частности сегментов.

Каждому разбиению в соответствие ставится свое значение оценки качества. Для оценки качества разбиения и соответствующего приближения изображения из N пикселей используется среднеквадратичное отклонение пикселей приближения от пикселей изображения или суммарная квадратичная ошибка $E=3N\sigma^2$, где коэффициент 3 учитывает число цветовых компонент в изображении.

Получение оптимальных приближений есть вычислительно сложная задача. Сложность экспоненциально возрастает с ростом числа пикселей и рассматриваемых значений числа кластеров от 1 до N. Оптимальные приближения можно аппроксимировать квази-оптимальными приближениями. Последние с достаточной скоростью генерируются в агломеративных и дивизимных алгоритмах минимизации ошибки E или среднеквадратичного отклонения σ [10], [11]. Поэтому, считаем разбиение изображения на кусочно-постоянные области оптимальным, если оно отвечает минимально возможному значению функционала качества.

При *сегментации* пиксели каждого вложенного изображения составляют единственный связный сегмент. При *кластеризации* пикселей *вложенное изображение* состоит из нескольких несмежных сегментов.

Идея кластеризации пикселей изображения классическим методом Уорда

Идея метода Уорда [2] заключается в последовательном слиянии пар кластеров в один. Оценкой качества служит приращение ΔE суммарной квадратичной ошибки E. При слиянии на каждом шаге выбирается пара кластеров i, j, объединение которых возвращает минимум приращения ΔE_{merge}:

$$\Delta E_{merge}(i,\ j) = \frac{n_i n_j}{n_i + n_j} \left\| I_i - I_j \right\|^2 \geq 0$$

где i, j – кластеры, n_i, n_j – число пикселей в кратерах, I_i, I_j - средние значения трехмерной яркости пикселей внутри кластеров.

Под приближением понимаем разбиение исходного изображения на однородные по цвету области, значения пикселей которых усреднены внутри кластеров. Последовательность приближений образует иерархию, так как очередное приближение получается из предыдущего объединением двух областей.

Достоинства метода Уорда:

1. Метод работает при любом числе кластеров от 1 до N числа элементов в исходном множестве, что удобно при разработке методов «слепой» сегментации.
2. Метод использует критерий качества, который позволяет оценить качество объединения пар и выбрать оптимальное решение из набора вариантов. Выбор на каждом шаге определяется минимальным значением функционала качества.
3. Метод позволяет проследить последовательность объединений пикселей в укрупненные множества. Запоминание иерархии позволяет двигаться в обратном направлении до необходимого состояния.
4. Метод Уорда как алгоритм кластеризации пикселей относится к адекватным методам. Это один из немногих методов, которые реально минимизирует суммарное квадратичное отклонение.

Недостатки метода Уорда:

1. Методу свойственна неоднозначность решений в зависимости от выбора пар для слияния при равнозначных вариантах.
2. Вычислительная сложность метода Уорда возрастает квадратично с ростом числа пикселей.

Первый недостаток незначителен и проявляется при независящих от исходного кода программы факторах, например, при переносе программного кода на платформу с иной операционной системой; при смене настроек работы процессора и проч. Второй недостаток кроется в самом программном коде и зависит от навыков программиста. Он затрудняет применение классического метода Уорда напрямую в обработке изображений, в частности в задачах сегментации, и требует нетривиальных решений.

В настоящей работе рассматривается блок-схема из комбинации базовых алгоритмов с иерархической структурой данных, позволяющих подготовить данные для последующей кластеризации классическим методом Уорда.

Типовая блок-схема последовательности алгоритмов

Обзор методов сегментации изображений [1] и способов повышения вычислительной эффективности позволил сформировать типовую схему последовательности алгоритмов, обеспечивающую снижение вычислительной сложности классического метода Уорда.

На рис. 1 представлена схема кластеризации пикселей или сегментации изображения, состоящая из трех последовательных этапов. Предлагаемая схема является типовой, поскольку первые два блока имеют различные варианты программной реализации. В первом блоке выполняется быстрое построение грубой иерархии. Во втором блоке выполняется улучшение качества заданного приближения при фиксированном числе цветов (кластеров). В третьем блоке выполняется кластеризация суперпикселей методом Уорда.

Блок «а) построение иерархии связных сегментов» быстро строит грубую иерархию приближений, генерация которой возможна двумя разными алгоритмами. *Приближением* изображения называется его кусочно-постоянное разбиение на фиксированное число кластеров. Блок «а)» имеет два различных варианта программной реализации. Реализация первого варианта заключается в применении модели Мамфорда-Шаха для построения грубой иерархии сегментов, посредством укрупнения сегментов на каждом шаге. Второй вариант реализации заключается в разделении изображения на фрагменты регулярной сеткой с последующей обработкой каждого фрагмента как самостоятельного изображения классическим методом Уорда с последующим объединением иерархий в одну.

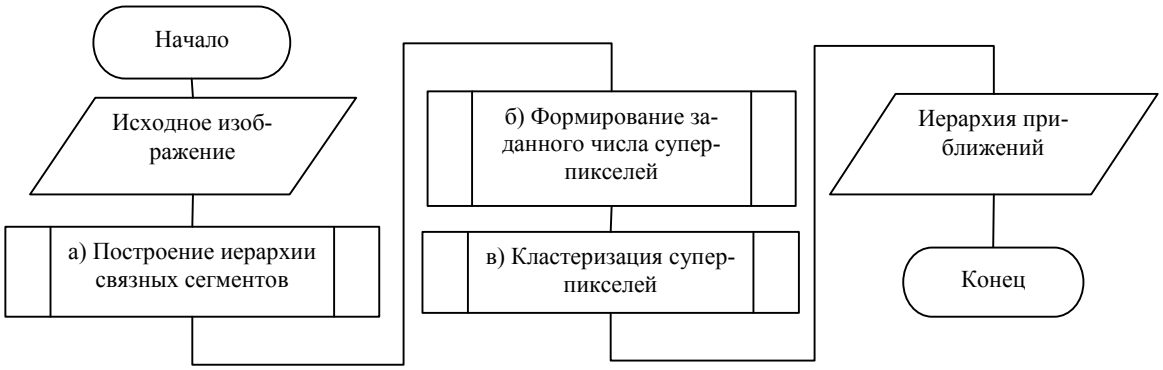


Рис. 1. Типовая блок-схема последовательности алгоритмов аппроксимации изображения

Блок «б) Формирование заданного числа суперпикселей» улучшает качество заданного приближения при фиксированном числе кластеров (цветов). Для этого блока разработаны два базовых алгоритма улучшения качества заданного разбиения изображения: *SI*-метод (Segmentation Improvement) [9], [12] и *K-meanless* метод (метод К-средних-без-средних) [13], [14]. Множество программных реализаций блока «б)» обусловлено как возможностью комбинированного сочетания пары базовых методов *SI* и *K-meanless* (отдельно, последовательно, циклично), так и версиями самих методов (сегментарная, кластерная).

Предлагаемый способ организации вычислительного процесса позволяет преодолеть чрезмерно высокую вычислительную сложность классического метода Уорда.

Применение обратимых операций в обработке изображений

Впервые идея применения обратимых операций в вычислительной технике предложена в работе [15]. В работе [9] описано применение встречной пары операций слияния и разделения в обработке изображений. Отличие развиваемой в настоящей работе идеи от оригинальной [15] в том, что обратная операция слияния по отношению к прямой операции разделения выполняется не обязательно с теми же элементами, что участвовали в прямой операции. Тем самым операция «отката в прошлое состояние» обобщена до операции «выхода в новое состояние».

Применение обратимых вычислений заключается в нахождении нового разбиения, характеризуемого меньшим значением суммарной квадратичной ошибки E. Рис. 2 иллюстрирует идею применения пары обратимых операций слияния и разделения множеств (кластеров, сегментов). На оси абсцисс (Ox) отложены целочисленные отсчеты чисел цветов (или кластеров) от 1 до N числа писклей в изображении. На оси ординат (Oy) откладывается значение приращения суммарной квадратичной ошибки, характеризующее качества разбиения изображения на множество (кластеров, сегментов) при каждом числе цветов (кластеров).

Кривая линия «1» соответствует процессу быстрого построения грубой иерархии сегментов (блок «а»). Такая иерархия сегментов образована последовательным объединением пикселей в сегменты, укрупнением сегментов слиянием смежных. Графически построение иерархии изображено движением справа налево от N числа сегментов до 1.

Кривая линия «2» показывает последовательное улучшение иерархии разбиением сегментов на составляющие (блок «б»). На каждом шаге выбирается сегмент, разделение которого сопровождается максимальным падением суммарной квадратичной ошибки E. Так, генерируется новое разбиение, характеризуемое меньшим значением E, то есть новое разбиение при том же количестве сегментов качественнее аналогичного другого. На рис. 2 кривая линия «2» получается отметкой на координатной плоскости оценок качеств разбиений при движении слева на право от 1 до 1000 кластера.

Кривая линия «3» соответствует результату блока «в) построения точной иерархии» классическим методом Уорда.

Методы улучшения качества заданного разбиения

SI-метод улучшения качества заданного разбиения изображения заключается в разделении одного сегмента на два и объединения пары смежных в один. На каждом шаге выбирается тройка (1, 2, 3) такая, чтобы по выполнении встречной пары операций «слияния-разделения» значение ошибки падала, соответственно, качество разбиения росло. Сегмент 1 выбираем из расчета максимального падения значения приращения ошибки при разделении его на составные части:

$$\min \Delta E_{divide}(1) < 0 \Leftrightarrow \max \Delta \left| E_{divide}(1) \right| < 0$$

Сегменты 2, 3 объединяем так, чтобы значение суммарной квадратичной ошибки минимально увеличилось:

$$\min \Delta E_{merge}(2,3) > 0.$$

В процессе выполнения алгоритма число сегментов изображения (кластеров пикселей) фиксировано. Показателем качества разбиения служит значение суммарной квадратичной ошибки E. При фиксированном числе кластеров пикселей или сегментов изображения лучшему разбиению соответствует меньшее значение E. *SI*-метод заключается в циклическом выполнении пары реверсивных операций слияния-разделения с множествами писклей или сегментами изображения при их фиксированном числе. На каждом шаге метод стремимся максимально понизить величину ско, характеризующую качество оптимизации разбиения:

$$\min \Delta E_{combine}(1,2,3) < 0 \Leftrightarrow \max \Delta \left| E_{combine}(1,2,3) \right| < 0$$

Условие останова состоит в том, как только суммарное приращение ошибки превысит 0:

$$\Delta E_{combine}(1,2,3) = \Delta E_{divide}(1) + \Delta E_{merge}(2,3) > 0 .$$

Существует две версии *SI*-метода: сегментарная и кластерная. В *сегментарной* версии *SI*-метода объединяемые множества (2, 3) обязательно должны быть смежными, в *кластерной* версии – любыми. На рис. 3 показана схема сегментарной версии *SI*-метода.

Метод «*K-meanless*» *(«K-средних-без-средних»)* реклассифицирует пиксели изображения согласно критерию:

$$\sqrt{\frac{n_2}{n_2+k}}\left\|I-I_2\right\| < \sqrt{\frac{n_1}{n_1-k}}\left\|I-I_1\right\|$$

Если опустить подкоренные выражения, то критерий совпадает с критерием реклассификации пикселей в классическом методе *K-means* [4]. В отличие от *K-means*, метод «К-средних-без-средних» [14] обходятся без самостоятельного вычисления средних по кластерам значений. Использование явных выражений для приращений суммарной квадратичной ошибки вместо вычислений самих значений функционала упрощает вычисления с массивами данных, т. к. позволяет обойтись без суммирования квадратов яркостей пикселей изображения.

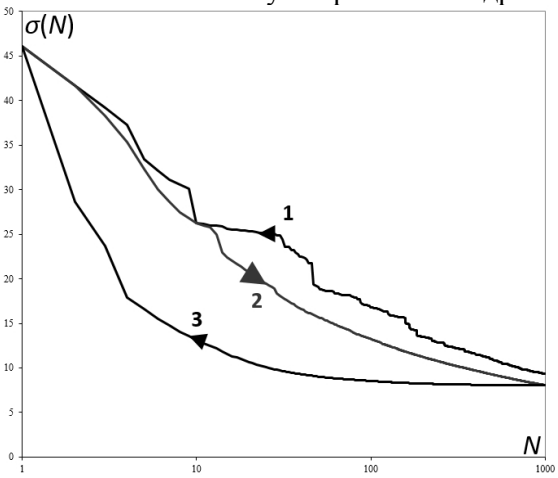


Рис. 2. Идея обратимых вычислений

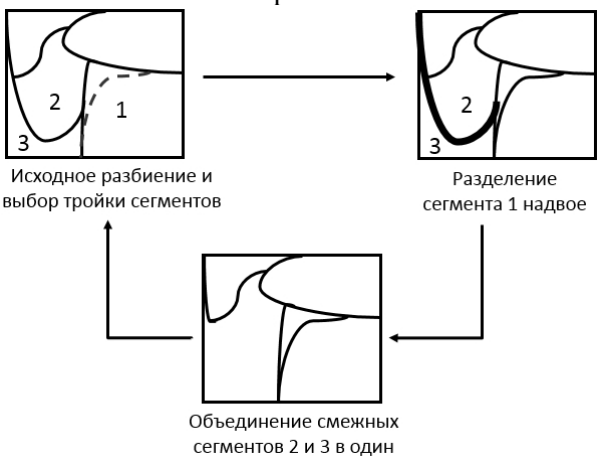


Рис. 3. Идея *SI*-метода

Экспериментальные результаты

Улучшение качества заданного приближения. На рис. 4 и 5 представлены результаты улучшений качеств заданных разбиений. Исходное изображение расположено на рис. 4а. Его исходное приближение на рис. 4б задано регулярной сеткой и получено сжатием в 64 раза с последующим восстановлением исходного размера. Каждая клетка представлена усреднением цветов, входящих в нее пикселей. Во втором ряду представлены результаты улучшения качества исходного приближения последовательным применением SI-метода (рис. 4в) и K-meanless метода (рис. 4г). Под каждым приближением приведены показатели качества - значения суммарных квадратичных отклонений σ.

На рис. 5в приведено «ошибочное» приближение, которое существенно отличается от заданного регулярной сеткой на рис. 4б. Для его генерации использовано вспомогательное изображение «Мандрил» рис. 5а такого же размера 512x512, которое было редуцировано до 1024 цветов/кластеров (рис. 5б). Затем полученные области вспомогательного разбиения (рис. 5б) заполнены пикселями исходного изображения «Лена» (рис. 4а). Конечное улучшенное приближение на рис. 5г получено применением пары *SI* и *K*-meanless методов в цикле.

Суммарная квадратичная ошибка, сгенерированного «ошибочного» приближения изображения «Лены» (рис. 5в) на треть выше (качество хуже), чем у приближения, полученного регулярной сеткой (рис. 4б). При этом, суммарное квадратичное отклонение итогового улучшенного приближения рис. 5г (σ=7,47459) меньше, чем у аналогичного на рис. 4г (σ=8,19082). Заметим, что применение пары методов (*SI* и *K*-meanless) в цикле возвращает более лучший результат, чем их разовое последовательное применение.

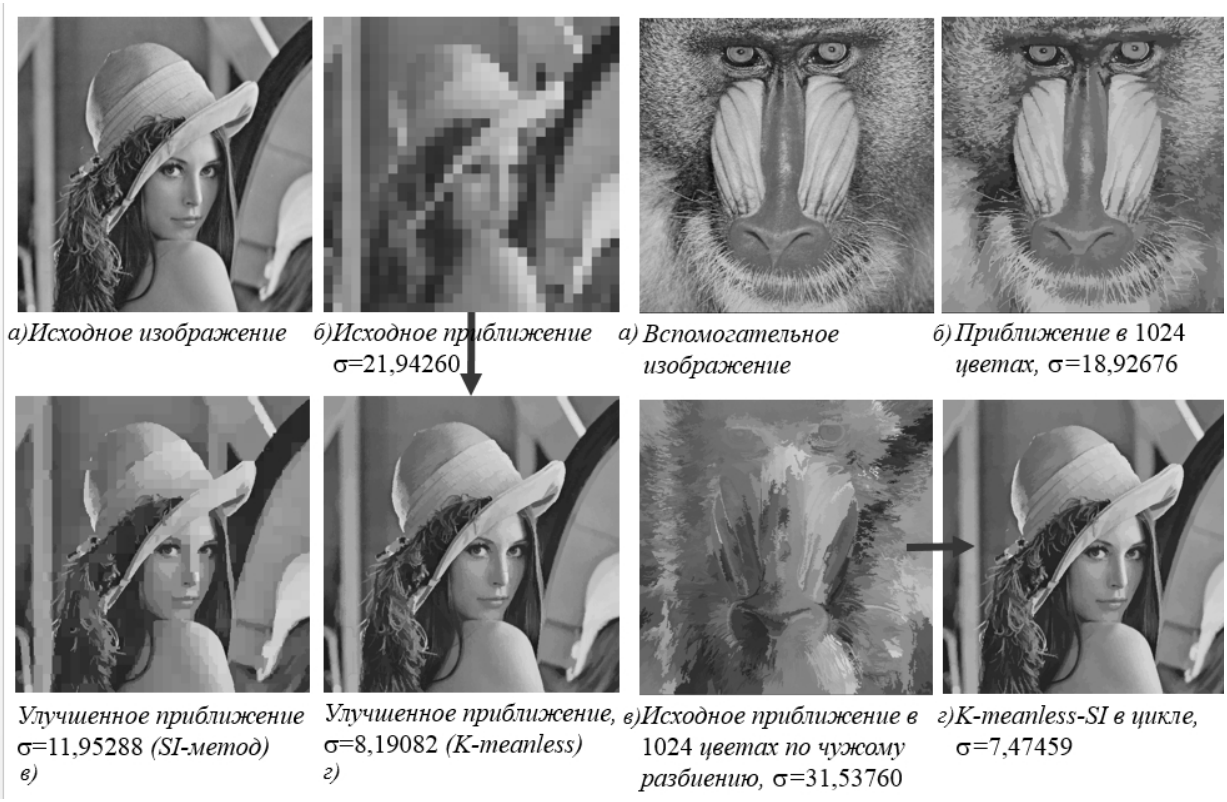


Рис. 4. Последовательное применение *SI* и *K-meanless* методов

Рис. 5. Циклическое применение *SI* и *K-meanless* методов

Улучшение качества иерархии приближений представлено на рис. 6. В первом ряду представлена конечная часть последовательности иерархической сегментации, где под каждым приближением приведена его оценка качества. Пример рассматривается справа налево от 6-го до 1-

го шага поскольку иерархия сегментов образуется объединением смежных пикселей, укрупнением сегментов/слиянием смежных.

Во второй строке рис. 6 приведена иерархия кластеризации пикселей. Каждое приближение которой поучена при помощи процедуры промежуточного улучшения качества (*SI*-метод, *K*-meanless). Улучшение качества наблюдаемо как субъективно визуально, так и объективно по меньшему значению суммарной квадратичной ошибки.

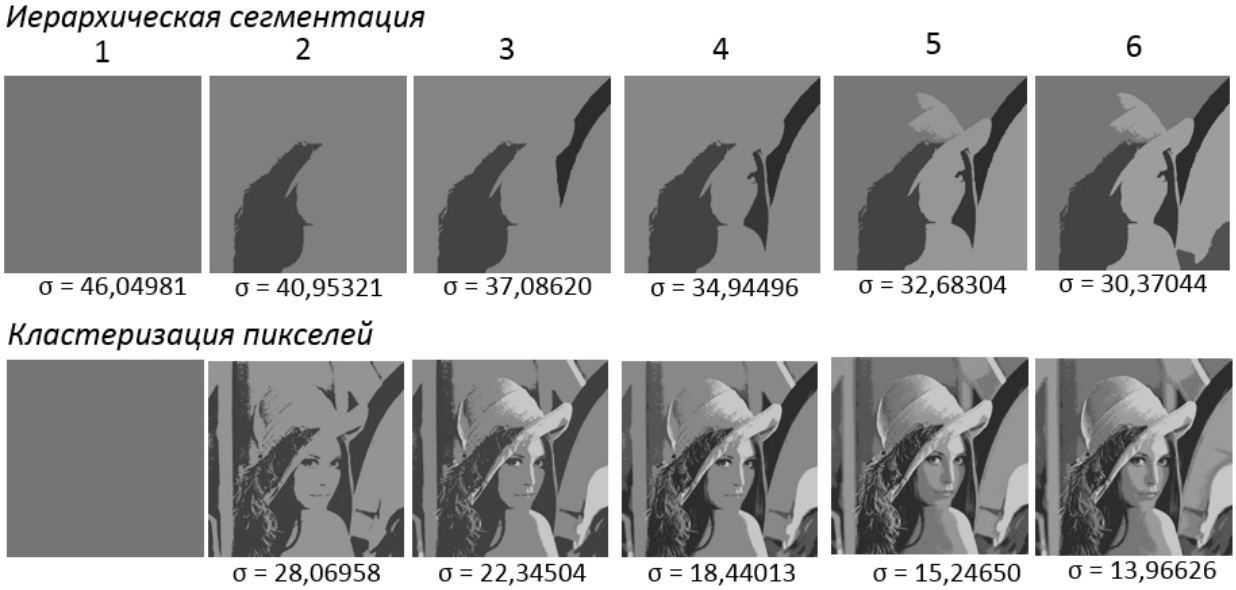


Рис. 6. Примеры последовательности иерархической сегментации и кластеризации пикселей

Выделение объектов на стереопаре. Понятие «одинаковое выделение» еще не формализовано, поэтому данный пример является эвристическим решаем поставленной задачи. На рис. 10 приведен пример иллюстрирующий выделение одинаковых объектов на различных изображения. Слева представлена исходные раздельные (рис. 7а) и совмещенные стереопары (рис. 7б). Справа (рис. 8а, 8б) - обработанные стереопары. Под каждым изображением приведен его показатель качества.

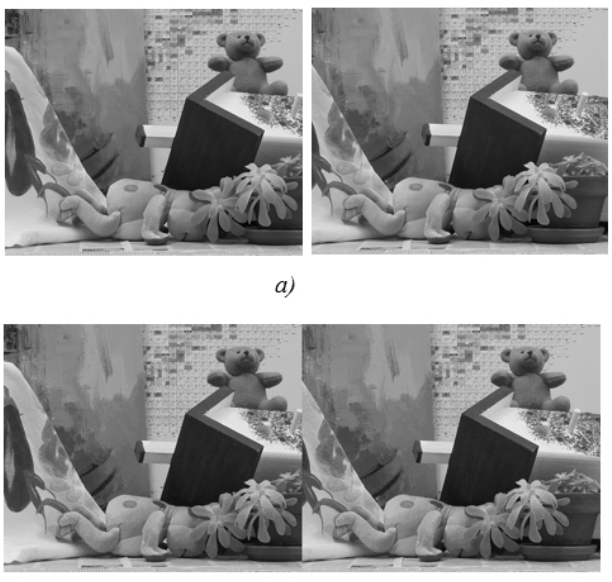


Рис. 7. Исходные стереопары

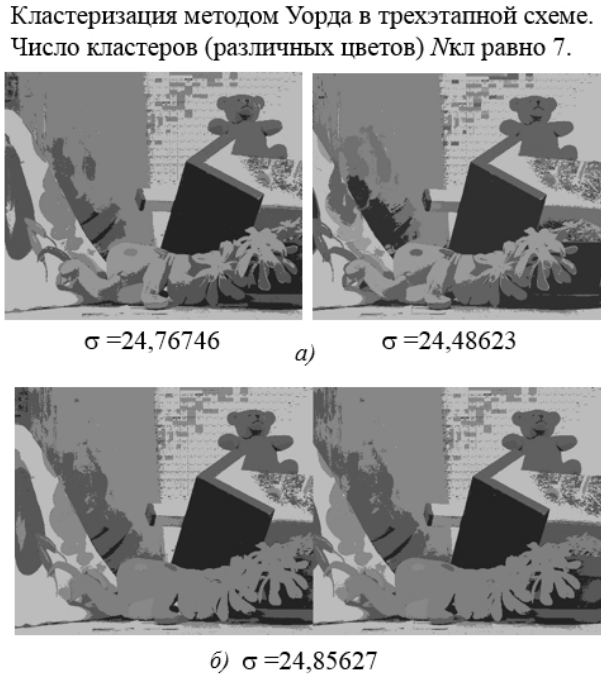


Рис. 8. Обработанные стереопары

На совмещенных стереопарах одинаковые тематические объекты: розовый мишка «Тедди», зеленый лягушонок, фоновые детали, – выделены одинаков. В то время как на раздельных снимание они отличаются.

Заключение

В работе приведены требования к актуальным алгоритмам сегментации изображений; обосновано предпочтение группы методов кластеризации данных;

1) сопоставлен ряд классически методов кластерного анализа: метод Оцу, Уорда, К-средних, модель Мамфорда-Шаха (Таблица 1);

2) обосновано преимущество метода Уорда перед остальными сопоставляемыми методами;

3) описана идея метода Уорда, его достоинства и недостатки;

4) приведены способы организации вычислительного процесса для преодоления вычислительной сложность классического метода Уорда;

5) описана типовая схема последовательности алгоритмов для скоростной аппроксимации изображения иерархической последовательностью приближений (рис. 1).

6) описано применение идеи обратимых операций в обработке изображений (рис. 2);

7) описаны методы улучшения качества заданного разбиения, реализуемые в блоке «б) формирование заданного числа суперпикселей» (рис. 3).

Предложенная трехэтапная схема позволяет:

1) обойти проблему вычислительной сложности за счет разделения процесса обработки на этапы;

2) строить различные программные реализации блоков при условии использования единой структуры данных.

Приведенные экспериментальные результаты свидетельствуют о пригодности рассматриваемой схемы к:

1) улучшению качества любой традиционной сегментации (рис. 4, 5);

2) улучшению качества иерархической последовательности приближений (рис. 6);

3) выделению одинаковых объектов на различных изображениях (рис. 7, 8).

ЛИТЕРАТУРА

1. Поршнев С. В., Левашкина А. О. Универсальная классификация алгоритмов сегментации изображений //Журнал научных публикаций аспирантов и докторантов. – 2008. – Т. 3. – С. 163-172.
2. Ward J.H., Jr. Hierarchical grouping to optimize an objective function. //J. Am. Stat. Assoc. 1963. Vol. 58, Issue 301, P. 236-244.
3. Otsu N. A threshold selection method from gray-level histograms //IEEE transactions on systems, man, and cybernetics. – 1979. – Т. 9. – №. 1. – С. 62-66.
4. Lloyd S. P. Least square quantization in PCM. Bell Telephone Laboratories Paper. Published in journal much later: Lloyd, SP: Least squares quantization in PCM //IEEE Trans. Inform. Theor.(1957/1982) Google Scholar. – 1957.
5. Mumford D., Shah J. Boundary detection by minimizing functionals // IEEE Conference on Computer Vision and Pattern Recognition. – 1985. – Т. 17. – С. 137-154.
6. Mumford D., Shah J. Optimal approximations by piecewise smooth functions and associated variational problems //Communications on pure and applied mathematics. – 1989. – Т. 42. – №. 5. – С. 577-685.
7. Liao P. S. et al. A fast algorithm for multilevel thresholding //J. Inf. Sci. Eng. – 2001. – Т. 17. – №. 5. – С. 713-727.
8. Wang Z., Bovik A. C. A universal image quality index //IEEE signal processing letters. – 2002. – Т. 9. – №. 3. – С. 81-84.
9. Харинов М. В., Ханыков И. Г. Оптимизация кусочно-постоянного приближения сегментированного изображения //Труды СПИИРАН. – 2015. – Т. 3. – №. 40. – С. 183-202.
10. *Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика: Классификация и снижение размерности // М.: Финансы и статистика, 1989. 607 с.
11. *Мандель И.Д.* Кластерный анализ // М.: Финансы и статистика. 1988. 176 с.
12. Khanykov I.G., Kharinov M.V., Patel C. Image Segmentation Improvement by Reversible Segment Merging. (2017, December 1-2). Int. Conf. on Soft Computing and its Engineering Applications, icSoftComp-2017, IEEE Gujarat Section Proceedings.
13. Kharinov M. Reclassification formula that provides to surpass K-means method //arXiv preprint arXiv:1209.6204. – 2012.
14. Dvoenko S. D. Meanless k-means as k-meanless clustering with the bi-partial approach //Proceedings of PRIP 2014 Conference, Minsk. – 2014.
15. Toffoli T. Reversible computing //International Colloquium on Automata, Languages, and Programming. – Springer, Berlin, Heidelberg, 1980. – С. 632-644.

I.G. Khanykov (Institute for Informatics and Automation of the Russian Academy of Sciences, Saint Petersburg, Russia)

The Technique for the Acceleration of the Classical Ward's Method for the Clustering of the Image Pixel

The classical methods of the cluster analysis that meet the actual requirements for modern image segmentation algorithms are compared. The classical methods of cluster analysis applicable in segmentation are compared: the methods of Ward, Otsu, K-averages, the Mumford-Shah model,. The choice of the classical Ward's method is justified as well as its advantages and disadvantages are presented. A typical block scheme of a sequence of algorithms is proposed, which allows to bypass the problem of computational complexity that is pecular to Ward's method. The application of the idea of the reversible operations in image processing is described. The experimental results on improving the quality of traditional segmentation improving as well as the quality of the hierarchical consequence of piece-wise constant partitions and identical object isolation in the different images are presented.

УЧЕТ ХАРАКТЕРА ПАРАМЕТРИЧЕСКОЙ НЕОРПРЕДЕЛЕННОСТИ ДИНАМИЧЕСКОЙ МОДЕЛИ ПРИ РАСЧЕТЕ РОБАСТНЫХ РЕГУЛЯТОРОВ ДЛЯ ОБЪЕКТОВ С САМОВЫРАВНИВАНИЕМ

Л. М. Яковис[1], **Н. А. Доронина**[2]
(*Санкт-Петербургский политехнический университет Петра Великого,
С.-Петербург, ул. Политехническая, д. 29, mc@spbstu.ru*)

Аннотация

Работа посвящена исследованию задачи настройки типовых пропорционально-интегральных (ПИ) и пропорционально-интегрально-дифференциальных (ПИД) регуляторов для устойчивых динамических объектов с инерционностью и запаздыванием в условиях, когда параметры модели точно не известны, и задана лишь область их возможных значений. Разработаны и опробованы с помощью средств имитационного компьютерного моделирования различные методы робастной настройки типовых регуляторов. Выполнен сравнительный анализ эффективности предложенных алгоритмов.

Ключевые слова: пропорционально-интегральный регулятор, пропорционально-интегрально-дифференциальный регулятор, инерционный объект, запаздывание.

Введение

Классическая теория автоматического управления рассматривает системы управления, для которых доступно точное математическое описание, однако такая ситуация является идеализированной. В реальных задачах неизбежно присутствует неопределенность, а потому используемое управление должно быть работоспособным при наличии неопределенности. Такое управление называется робастным.

Если модель описывает физический объект (механический, электрический, экономический и т.п.), то во многих случаях его параметры могут неконтролируемым образом меняться в процессе эксплуатации. При этом структура соотношений, описывающих динамику объекта, может быть известной и неизменной. В таких ситуациях можно говорить о параметрической неопределенности.

Разработка эффективных методов робастной настройки регуляторов для систем управления с обратной связью одна из актуальных задач теории и практики автоматического управления. Трудности решения этой задачи применительно к регуляторам с заданной структурой передаточной функции (к этому классу относятся широко распространенные в промышленности типовые регуляторы) рассмотрены в [1]. Задача дополнительно усложняется для объектов с запаздыванием по каналу передачи управляющих воздействий [2].

В данной работе рассматривается как интервальная неопределенность, когда известны лишь интервалы возможных значений параметров математической модели объекта управления, так и неопределенность не интервального типа (будем называть ее неинтервальной неопределенностью), когда параметры модели объекта могут оказаться в любой точке заданной области произвольного вида. Исследуются задачи настройки типовых пропорционально-интегральных (ПИ) и пропорционально-интегрально-дифференциальных (ПИД) регуляторов для устойчивых динамических объектов с инерционностью и запаздыванием в условиях, когда параметры модели точно не известны, но задана область их возможных значений.

Робастная настройка типовых регуляторов при интервальной неопределенности. Вначале рассмотрим задачу робастной настройки типовых регуляторов в условиях интервальной неопределенности. В качестве базовой системы возьмем инерционный объект первого порядка с запаздыванием, показанный на рис. 1. Рассматривая систему в отклонениях от режимных значений, функционирующую в условиях возмущений n, поставим задачу стабилизации выходной переменной $y(t)$ в окрестности нулевого значения с использованием типового регулятора.

[1] Доктор технических наук, старший научный сотрудник.
[2] Магистрант

В варианте интервальной неопределенности параметры объекта управления: коэффициент усиления k, постоянная времени T и запаздывание τ, известны с точностью до интервалов их возможных значений, то есть

$$\underline{k} \le k \le \overline{k},\ \underline{T} \le T \le \overline{T},\ \underline{\tau} \le \tau \le \overline{\tau}. \qquad (1)$$

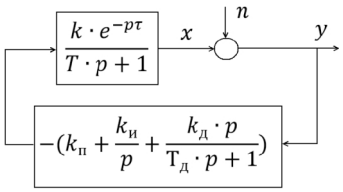


Рис. 1. Инерционный объект первого порядка с запаздыванием, управляемый ПИД-регулятором

Необходимо найти настройки параметров ПИ или ПИД-регулятора, которые обеспечивают работоспособность замкнутой системы в любой точке области возможных значений параметров объекта управления $\boldsymbol{G}$. В рассматриваемом варианте эта область представляет собой параллелепипед в пространстве (k, T, τ), показанный на рис. 2.

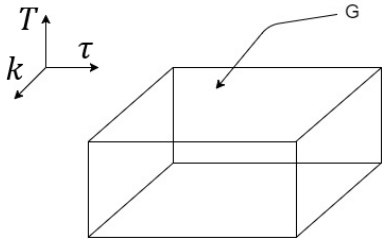


Рис. 2. Область интервальной параметрической неопределенности

Введем вектор параметров «истинной» модели объекта управления $\boldsymbol{Z} = [k, T, \tau]^T$ и вектор $\boldsymbol{A}$ параметров регулятора. Для ПИ-регулятора $\boldsymbol{A} = [k_p, k_i]^T$, для ПИД-регулятора $\boldsymbol{A} == [k_p, k_i, k_d]^T$ (малая постоянная времени T_d «реального» ПИД-регулятора обычно берется в заданной пропорции к остальным настраиваемым параметрам и не включается в состав вектора варьируемых параметров $\boldsymbol{A}$). При традиционном подходе к расчету регуляторов в рамках одной из известных методик [3] $\boldsymbol{A}$ рассчитывается в зависимости от параметров расчетной модели управляемого объекта $\boldsymbol{Z}_r = [k_r, T_r, \tau_r]^T$, то есть $\boldsymbol{A} = \boldsymbol{F}(\boldsymbol{Z}_r)$, причем в условиях неопределенности $\boldsymbol{Z}_r$ может отличаться от $\boldsymbol{Z}$.

Пусть показателем качества управления является какой-либо интегральный показатель, например, интегральный квадратичный критерий (ИКК)

$$J = \int_0^{\infty} [y(t)]^2 dt,$$

характеризующий сходимость выходной переменной к заданному значению при подаче ступенчатого возмущения n. Значение J определяется как параметрами объекта, так и параметрами регулятора, поэтому можно написать $J = \varphi(k, T, \tau, k_{\text{п}}, k_{\text{и}}, k_{\text{д}})$.

Исходя из стремления обеспечить лучшее качество управления в худшем из возможных вариантов, робастные настройки $k_{\text{п}}^{\text{роб}}, k_{\text{и}}^{\text{роб}}, k_{\text{д}}^{\text{роб}}$ находятся в результате решения задачи определения

$$J_{\min max} = \min_{k_{\text{п}}, k_{\text{и}}, k_{\text{д}}} \Psi(k_{\text{п}}, k_{\text{и}}, k_{\text{д}}), \qquad (2)$$

где

$$\Psi(k_{\text{п}}, k_{\text{и}}, k_{\text{д}}) = \max_{(k,T,\tau) \in G} \varphi(k, T, \tau, k_{\text{п}}, k_{\text{и}}, k_{\text{д}}). \qquad (3)$$

Ввиду сложности аналитического решения задачи (2), (3) [1] можно предложить приближенные решения. Один из способов базируется на следующих предположениях.

В соответствии с компенсационным методом [4] при известной модели объекта k, T, τ субоптимальный ПИ-регулятор рассчитывается по формулам

$$\tilde{k}_{\text{п}} = \frac{a \cdot T}{k \cdot \tau}, \qquad \tilde{k}_{\text{и}} = \frac{a}{k \cdot \tau},$$

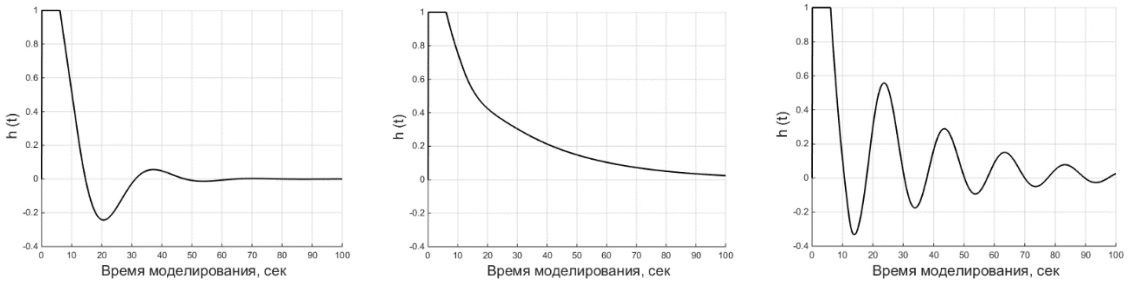
где константа a зависит от критерия качества управления.

Для ПИД-регулятора компенсационный метод дает

$$\tilde{k}_{\text{п}} = \frac{a \cdot T}{k \cdot \tau} + \frac{b}{k}, \qquad \tilde{k}_{\text{и}} = \frac{a}{k \cdot \tau}, \qquad \tilde{k}_{\text{д}} = \frac{b \cdot T}{k},$$

где константы a и b также зависят от критерия качества управления.

Анализ большого числа имитационных экспериментов, моделирующих поведение оптимально (или субоптимально) настроенной замкнутой системы, а также – поведения системы управления при ее разладке (то есть при несоответствии настроек регулятора оптимальным или субоптимальным значениям) показал [5], что при отработке приведенных к выходу ступенчатых возмущений переходные процессы могут быть трех типов (см. рис. 3).



a) b) c)

Рис. 3. Кривые переходных процессов при: a) оптимальных настройках регулятора, b) «слабых» настройках регулятора, c) «сильных» настройка регулятора

Опасность потери устойчивости возникает при слишком сильных настройках, т.е. когда $k_{\text{п}} > \tilde{k}_{\text{п}}$, $k_{\text{и}} > \tilde{k}_{\text{и}}$, $k_{\text{д}} > \tilde{k}_{\text{д}}$. Поэтому желательно, чтобы для всех точек зоны неопределенности ***G*** выполнялись условия

$$k_{\text{п}} \le \tilde{k}_{\text{п}}, \qquad k_{\text{и}} \le \tilde{k}_{\text{и}}, \qquad k_{\text{д}} \le \tilde{k}_{\text{д}}. \qquad (4)$$

Ясно, что если имеет место неопределенность интервального типа, то условия (4) будут выполнены, только если для ПИ-регулятора

$$k_{\text{п}}^{\text{роб}} \le \min_{k,T,\tau \in G} \tilde{k}_{\text{п}} = \frac{a \cdot \underline{T}}{\overline{k} \cdot \overline{\tau}}, \qquad k_{\text{и}}^{\text{роб}} \le \min_{k,T,\tau \in G} \tilde{k}_{\text{и}} = \frac{a}{\overline{k} \cdot \overline{\tau}}, \qquad (5)$$

а для ПИД-регулятора

$$k_{\text{п}}^{\text{роб}} \le \min_{k,T,\tau \in G} \tilde{k}_{\text{п}} = \frac{a \cdot \underline{T}}{\overline{k} \cdot \overline{\tau}} + \frac{b}{\overline{k}}, \qquad k_{\text{и}}^{\text{роб}} \le \min_{k,T,\tau \in G} \tilde{k}_{\text{и}} = \frac{a}{\overline{k} \cdot \overline{\tau}}, \qquad k_{\text{д}}^{\text{роб}} \le \min_{k,T,\tau \in G} \tilde{k}_{\text{д}} = \frac{b \cdot \underline{T}}{\overline{k}}. \quad (6)$$

Выполнение условий (4) предохраняет от слишком сильных настроек регулятора, чреватых потерей устойчивости. Слишком слабые настройки тоже связаны с потерями качества управления, поэтому в рамках (4) следует выбрать такие настройки $k_{\text{п}}, k_{\text{и}}, k_{\text{д}}$, которые будут максимально близки к субоптимальным значениям $\tilde{k}_{\text{п}}, \tilde{k}_{\text{и}}, \tilde{k}_{\text{д}}$. Для формализации условия близости можно ввести различные показатели, например

$$I = (\delta k_{\text{п}})^2 + (\delta k_{\text{и}})^2 + (\delta k_{\text{д}})^2,$$

где

$$\delta k_{\text{п}} = \frac{k_{\text{п}} - \tilde{k}_{\text{п}}}{\tilde{k}_{\text{п}}}, \qquad \delta k_{\text{и}} = \frac{k_{\text{и}} - \tilde{k}_{\text{и}}}{\tilde{k}_{\text{и}}}, \qquad \delta k_{\text{д}} = \frac{k_{\text{д}} - \tilde{k}_{\text{д}}}{\tilde{k}_{\text{д}}}.$$

Заметим, что I зависит как от характеристик объекта управления k, T, τ, так и от настроек регулятора $k_{\text{п}}, k_{\text{и}}, k_{\text{д}}$, т.е. представляет собой некоторую функцию от перечисленных показателей $I = S(k, T, \tau, k_{\text{п}}, k_{\text{и}}, k_{\text{д}})$.

Учитывая, что правила настройки регуляторов при известных параметрах объекта управления есть некоторые функции от этих параметров, можно написать:

$$\tilde{k}_{\text{п}} = \varphi_{\text{п}}(k,T,\tau), \qquad \tilde{k}_{\text{и}} = \varphi_{\text{и}}(k,T,\tau), \qquad \tilde{k}_{\text{д}} = \varphi_{\text{д}}(k,T,\tau).$$

Теперь можно сформулировать задачу определения робастных настроек. Пусть

$$\vartheta(k_{\text{п}},k_{\text{и}},k_{\text{д}}) = \max_{(k,T,\tau)\in G}\left\{ S(k,T,\tau,k_{\text{п}},k_{\text{и}},k_{\text{д}}) \middle| \begin{matrix} k_{\text{п}} \le \varphi_{\text{п}}(k,T,\tau), \\ k_{\text{и}} \le \varphi_{\text{и}}(k,T,\tau), \\ k_{\text{д}} \le \varphi_{\text{д}}(k,T,\tau). \end{matrix} \right\} \qquad (7)$$

Тогда робастные настройки находятся решением задачи определения

$$I_{\min max} = \min_{k_{\text{п}},k_{\text{и}},k_{\text{д}}}\left\{ \vartheta(k_{\text{п}},k_{\text{и}},k_{\text{д}}) \middle| \begin{matrix} k_{\text{п}} \le \varphi_{\text{п}}(k,T,\tau), \\ k_{\text{и}} \le \varphi_{\text{и}}(k,T,\tau), \\ k_{\text{д}} \le \varphi_{\text{д}}(k,T,\tau). \end{matrix} \right\} \qquad (8)$$

Применяя эту общую схему к варианту интервальной неопределенности типа (1), получим, что решение задачи (7), (8) находится по правилу

$$A_{rob} = F(Z_r), \text{где } Z_r = Z_{rob} = [\overline{k},\underline{T},\overline{\tau}]^T, \qquad (9)$$

то есть лежит на границах неравенств (5), (6). Отсюда следует, что для ПИ-регулятора робастные настройки рассчитываются по формулам

$$k_{\text{п}}^{\text{роб}} = \underline{\tilde{k}_{\text{п}}} = \frac{a\cdot\underline{T}}{\overline{k}\cdot\overline{\tau}}, \qquad k_{\text{и}}^{\text{роб}} = \underline{\tilde{k}_{\text{и}}} = \frac{a}{\overline{k}\cdot\overline{\tau}}. \qquad (10)$$

Подобным же образом рассматривается случай ПИД-регулятора

$$k_{\text{п}}^{\text{роб}} = \underline{\tilde{k}_{\text{п}}} = \frac{a\cdot\underline{T}}{\overline{k}\cdot\overline{\tau}} + \frac{b}{\overline{k}}, \qquad k_{\text{и}}^{\text{роб}} = \underline{\tilde{k}_{\text{и}}} = \frac{a}{\overline{k}\cdot\overline{\tau}}, \qquad k_{\text{д}}^{\text{роб}} = \underline{\tilde{k}_{\text{д}}} = \frac{b\cdot\underline{T}}{\overline{k}}. \qquad (11)$$

Заметим, что соотношения (10) и (11) совпадают с формулами, полученными ранее эвристическим путем [6].

Робастные настройки при неинтервальной неопределенности. Задание зоны неопределенности в виде интервалов возможных значений параметров модели управляемого объекта предполагает всевозможные сочетания значений параметров из этих интервалов. Геометрически, применительно к модели типа инерционного звена первого порядка с запаздыванием, это означает, что в пространстве (k,T,τ) точки, отвечающие возможным объектам, заполняют параллелепипед, соответствующий границам зоны неопределенности (см. рис. 2). Вместе с тем часто встречается ситуация, когда в силу взаимозависимости между характеристиками объекта управления они могут заполнять лишь определенную область внутри параллелепипеда (см. рис. 4). Учет этого факта позволяет сузить зону неопределенности и «усилить» робастные настройки, что, в свою очередь, может позволить улучшить средние показатели качества управления.

В случае, если область неопределенности представляет собой произвольную область $\boldsymbol{G}$, можно попытаться свести задачу расчета робастных настроек регулятора к интервальной неопределенности, применив так называемый *метод окаймляющего параллелепипеда* (МОП). Для этого сформируем окаймляющий параллелепипед $\boldsymbol{G'}$, т.е. параллелепипед минимального объема, содержащий область $\boldsymbol{G}$.

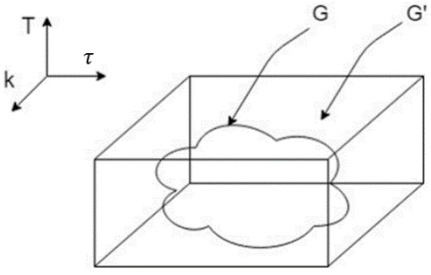


Рис. 4. Область неинтервальной неопределенности *G* и окаймляющий ее параллелепипед *G'*

Границы такого параллелепипеда находятся в результате решения 6 вспомогательных задач определения

$$\underline{k} = min\{k|(k,T,\tau) \in G\}, \qquad \overline{k} = max\{k|(k,T,\tau) \in G\},$$
$$\underline{T} = min\{T|(k,T,\tau) \in G\}, \qquad \overline{T} = max\{T|(k,T,\tau) \in G\},$$
$$\underline{\tau} = min\{\tau|(k,T,\tau) \in G\}, \qquad \overline{\tau} = max\{\tau|(k,T,\tau) \in G\}.$$

Окаймляющий параллелепипед $\boldsymbol{G'}$ должен иметь минимальный объем, т.к. излишнее занижение параметров настройки регулятора (см. рис. 3b) приводит к потере качества управления.

Решив далее задачу робастного управления для окаймляющего область $\boldsymbol{G}$ параллелепипеда $\boldsymbol{G'}$, мы получим регулятор, удовлетворяющий ограничениям (4) в любой точке области $\boldsymbol{G'}$, а значит и в любой точке вписанной в $\boldsymbol{G'}$ области $\boldsymbol{G}$.

Метод окаймляющего параллелепипеда имеет, однако, существенный недостаток: в ряде случаев его использование может приводить к слишком слабым регуляторам, что сопряжено с потерями качества управления. Одним из способов, направленных на устранение этого недостатка, является применение *метода аппроксимирующих параллелепипедов* (МАП). Рассмотрим этот метод, применительно к инерционному объекту с запаздыванием, который задается тремя параметрами (k,T,τ).

Пусть неопределенность параметров объекта управления задана областью $\boldsymbol{G}$ в трехмерном пространстве (k,T,τ).

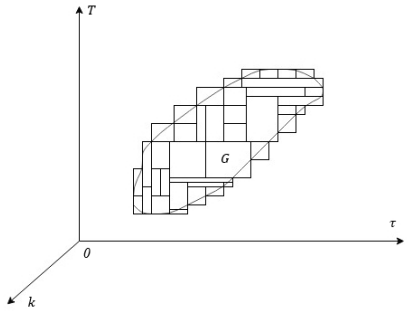


Рис. 5. Аппроксимация области $\boldsymbol{G}$ непересекающимися параллелепипедами

Допустим, что область $\boldsymbol{G}$ аппроксимирована N непересекающимся параллелепипедами $\widetilde{\boldsymbol{G}}_j$ в пространстве (k,T,τ), как это показано на рис. 5 (о способах такой аппроксимации будет сказано позже). Это значит, что каждая точка (k,T,τ), принадлежащая $\boldsymbol{G}$, попадает хотя бы в один параллелепипед[1], и, наоборот, в каждый аппроксимирующий параллелепипед попадает хотя бы одна точка из $\boldsymbol{G}$. Тогда каждый из параллелепипедов $\widetilde{\boldsymbol{G}}_j$ задается границами трех интервалов: $\underline{k}_j \le k_j \le \overline{k}_j,\ \underline{T}_j \le T_j \le \overline{T}_j,\ \underline{\tau}_j \le \tau_j \le \overline{\tau}_j,\ j = \overline{1,N}.$

Суть метода состоит в том, чтобы найти ограничение сверху на коэффициенты регулятора, которые гарантируют достаточный запас устойчивости для всех объектов из зоны неопределенности, и пытаться подойти как можно ближе к границам этой допустимой области настроек (чем максимально усилить робастные настройки). Реализация этих правил приводит к следующим алгоритмам робастной настройки:

a) для ПИ-регулятора

$$k_{\text{п}}^{\text{роб}} = \min_{j=\overline{1,N}}\left\{k_{\text{п}_j}^{\text{роб}}\right\} = \min_{j=\overline{1,N}}\left\{\frac{a\cdot\underline{T}_j}{\overline{k}_j\cdot\overline{\tau}_j}\right\},\ k_{\text{и}}^{\text{роб}} = \min_{j=\overline{1,N}}\left\{k_{\tau_j}^{\text{роб}}\right\} = \min_{j=\overline{1,N}}\left\{\frac{a}{\overline{k}_j\cdot\overline{\tau}_j}\right\},$$

b) для ПИД-регулятора

$$k_{\text{п}}^{\text{роб}} = \min_{j=\overline{1,N}}\left\{\frac{a\cdot\underline{T}_j}{\overline{k}_j\cdot\overline{\tau}_j} + \frac{b}{\overline{k}_j}\right\},\ k_{\text{и}}^{\text{роб}} = \min_{j=\overline{1,N}}\left\{\frac{a}{\overline{k}_j\cdot\overline{\tau}_j}\right\},\ k_{\text{д}}^{\text{роб}} = \min_{j=\overline{1,N}}\left\{\frac{b\cdot\underline{T}_j}{\overline{k}_j}\right\}.$$

Аналитическое задание области неопределенности $\boldsymbol{G}$. По-прежнему будем считать известными границы параллелепипеда возможных параметров объекта управления (1), но вместе

[1] Некоторые точки из **G** могут попасть на грани или в вершины примыкающих друг к другу параллелепипедов – такие точки принадлежат двум и более параллелепипедам.

с тем будем предполагать, что известны зависимости между параметрами, которые заданы системой равенств и неравенств

$$\boldsymbol{F_1}(k,T,\tau) = 0, \qquad \boldsymbol{F_2}(k,T,\tau) \le 0. \qquad (11)$$

Здесь $\boldsymbol{F_1}(k,T,\tau)$ и $\boldsymbol{F_2}(k,T,\tau)$ – известные вектор-функции. Рассмотрим далее процедуру аппроксимации G совокупностью параллелепипедов.

a) Отрезки $[\underline{k},\overline{k}], [\underline{T},\overline{T}], [\underline{\tau},\overline{\tau}]$ делятся на N_k, N_T, N_τ частей соответственно. Таким образом образуется сетка в трехмерном пространстве (k,T,τ), содержащая $N_k \times N_T \times N_\tau = N$ параллелепипедов. Чем гуще сетка, т.е. чем больше N_k, N_T, N_τ, тем точнее будет аппроксимация области G параллелепипедами.

b) Каждый из N параллелепипедов так же, как это описано в п.a), представляется в виде совокупности $M_k \times M_T \times M_\tau = M$ параллелепипедов с помощью сетки, имеющей $(M_k + 1) \times$ $\times (M_T + 1) \times (M_\tau + 1) = \widetilde{M}$ узлов.

c) Для каждого из $\widetilde{M}$ узлов каждого из N составляющих параллелепипедов делается проверка наличия у него общих точек с областью $\boldsymbol{G}$, задаваемой соотношениями (11), т.е. делается проверка выполнения условий

$$\boldsymbol{F_1}(k_j,T_j,\tau_j) = 0, \qquad \boldsymbol{F_2}(k_j,T_j,\tau_j) \le 0, \qquad (12)$$

где k_j, T_j, τ_j – координаты $j-$го узла, $j = \overline{1,\widetilde{M}}$.

Если соотношения (12) выполняются хотя бы для одного узла сетки, то делается вывод, что этот параллелепипед имеет общие точки с $\boldsymbol{G}$, а стало быть, он включается в число аппроксимирующих $\boldsymbol{G}$ параллелепипедов. В противном случае считается, что данный параллелепипед не имеет общих точек с $\boldsymbol{G}$, и он не включается в число аппроксимирующих.

По итогам такой проверки всех N параллелепипедов выявляется $\widetilde{N} < N$ параллелепипедов, образующих область $\widetilde{\boldsymbol{G}}$, которая аппроксимирует область $\boldsymbol{G}$.

Теперь задача робастной настройки ПИ и ПИД-регулятора для объекта управления $H(p) = \frac{k \cdot e^{-p\cdot\tau}}{T\cdot p+1}$ с зоной неопределенности параметров (11), внутренней по отношению к охватывающему ее параллелепипеду (1), может быть решена методом аппроксимирующих параллелепипедов применительно к области $\widetilde{\boldsymbol{G}}$.

Точечное задание зоны неопределенности $\boldsymbol{G}$. Наряду с МАП для приближенного описания зоны неопределенности (11) может применяться *метод точечной аппроксимации* (МТА). Определив границы окаймляющего область $\boldsymbol{G}$ параллелепипеда, его покрывают сеткой точек, а затем отбирают те из них, $(k_1,T_1,\tau_1),\dots,(k_L,T_L,\tau_L)$, для которых выполняются условия (11). Применяя идеологию, сходную с методом аппроксимирующих параллелепипедов, получим формулы для ПИ-регулятора

$$k_{\text{п}}^{\text{роб}} = \min_{j=\overline{1,L}}\left\{\tilde{k}_{\text{п}_j}\right\} = \min_{j=\overline{1,L}}\left\{\frac{a\cdot T_j}{k_j\cdot\tau_j}\right\}, \qquad k_{\text{и}}^{\text{роб}} = \min_{j=\overline{1,L}}\left\{\tilde{k}_{\text{и}_j}\right\} = \min_{j=\overline{1,L}}\left\{\frac{a}{k_j\cdot\tau_j}\right\}.$$

Подобным образом могут быть сформированы соотношения и для настройки робастных ПИД-регуляторов. Ясно, что неограниченно уменьшая шаг сетки при делении на части окаймляющего параллелепипеда, можно неограниченно увеличивать L, добиваясь все более точной аппроксимации области $\boldsymbol{G}$ множеством точек. Это, однако, может существенно увеличить время расчета робастных настроек. Важно также понимать, что точки при любом их количестве аппроксимируют $\boldsymbol{G}$ «изнутри» в отличие от параллелепипедов, аппроксимирующих $\boldsymbol{G}$ «снаружи». По этой причине МАП с большей вероятностью, чем МТА, обеспечит сохранение качества управления и, в частности, устойчивость замкнутой системы.

Распространенная ситуация заключается в том, что в распоряжении разработчиков системы управления имеются данные экспериментов по идентификации модели интересующего объекта управления или сходных с ним объектов. Например, это могут быть данные экспериментов по снятию разгонных характеристик. По данным экспериментов методами идентификации моде-

лей объектов управления (ОУ) могут быть определены параметры k,T,τ. В пространстве k,T,τ данные идентификации можно адекватно изобразить множеством точек (см. рис. 6)

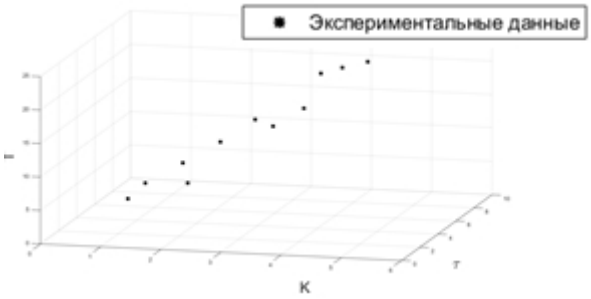


Рис. 6. Данные идентификации модели объекта управление в пространстве k,T,τ

При этом расчет робастных настроек может естественным образом производиться на основе МТА. Вместе с тем при малом объеме экспериментальных данных использовать точечную аппроксимацию не рекомендуется, т.к. аппроксимирующее множество $\widetilde{\boldsymbol{G}}$ содержит лишь малую выборку точек реальной зоны неопределенности $\boldsymbol{G}$. Смягчить данный недостаток, т.е. слабую представительность малой выборки, состоящей из небольшого числа взятых из $\boldsymbol{G}$ точек, может рассмотренный ранее метод аппроксимирующих параллелепипедов.

Вероятностно-статический подход к расчету робастных настроек. Применение метода Монте-Карло для повышения качества работы робастного управления. Если имеется малая экспериментальная выборка, то для получения более качественных робастных настроек возникает стремление увеличить объем выборки, дополнив ее точками пространства (k,T,τ), с высокой вероятностью принадлежащими множеству $\boldsymbol{G}$. При этом можно действовать следующим образом.

Введем вектор параметров модели объекта управления $\boldsymbol{X} = [k,T,\tau]^T$. Статистической обработкой данных $\boldsymbol{x_1},\dots,\boldsymbol{x_L}$ можно найти оценки математического ожидания $\widehat{M}(\boldsymbol{X})$ и матрицы коэффициентов парной корреляции $\widehat{\boldsymbol{K}}$.

Рассмотрим модель неопределенности параметров объекта управления

$$\boldsymbol{X} = \mathring{\boldsymbol{X}} + M(\boldsymbol{X}), \qquad \mathring{\boldsymbol{X}} = \boldsymbol{A}\cdot\boldsymbol{\vartheta}, \tag{13}$$

где

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \qquad \boldsymbol{\vartheta} = [\vartheta_1 \quad \vartheta_2 \quad \vartheta_3\ ]^T.$$

Здесь $\boldsymbol{A}$ – матрица подлежащих определению параметров модели (13), $\boldsymbol{\vartheta}$ – вектор не коррелированных друг с другом центрированных белых шумов с единичным средне- квадратичным отклонением.

Пусть $\boldsymbol{K} = M\left\{\left(\boldsymbol{X}-M(\boldsymbol{X})\right)\left(\boldsymbol{X}-M(\boldsymbol{X})\right)^T\right\} = \left\{\mathring{\boldsymbol{X}}\mathring{\boldsymbol{X}}^T\right\}$ - корреляционная матрица случайного вектора $\boldsymbol{X}$. Найдем такую матрицу $\boldsymbol{A}$, чтобы выполнялось условие

$$M\left\{\mathring{\boldsymbol{X}}\mathring{\boldsymbol{X}}^T\right\} = \boldsymbol{K}. \tag{14}$$

Подставляя (13) в левую часть (14), получим

$$M\left\{\mathring{\boldsymbol{X}}\mathring{\boldsymbol{X}}^T\right\} = M\{\boldsymbol{A}\boldsymbol{\vartheta}\boldsymbol{\vartheta}^{\mathbf{T}}\boldsymbol{A}^{\mathbf{T}}\} = \boldsymbol{A}\mathrm{M}\{\boldsymbol{\vartheta}\boldsymbol{\vartheta}^{\mathbf{T}}\}\boldsymbol{A}^{\mathbf{T}} = \boldsymbol{A}\boldsymbol{A}^{\mathbf{T}}, \qquad \left(\mathrm{M}\{\boldsymbol{\vartheta}\boldsymbol{\vartheta}^{\mathbf{T}}\} = \boldsymbol{E}\right).$$

Матрица $\boldsymbol{A}$ определяется в ходе решения матричного уравнения $\boldsymbol{A}\boldsymbol{A}^{\mathbf{T}} = \boldsymbol{K}$ (задача факторизации матрицы $\boldsymbol{K}$). Эта система квадратных матричных у равнений достаточно просто решается аналитически благодаря диагональной структуре матрицы $\boldsymbol{A}$.

Если при расчете $\boldsymbol{A}$ подставить в правую часть (14) оценку $\widehat{\boldsymbol{K}}$, то получим модель случайного процесса с требуемой корреляционной матрицей. Генерируя с помощью датчика случайных чисел вектора $\boldsymbol{\vartheta}$, подвергая их матричному преобразованию $\boldsymbol{A}$, и смещая полученные значения $\mathring{\boldsymbol{X}}$ в соответствии с (13) на величину $\widehat{M}(\boldsymbol{X})$, можно таким образом генерировать в любом коли-

честве набор случайных чисел, вероятностные характеристики которых согласованы с экспериментальными данными.

Применение многомерного нормального распределения. Второй подход для повышения качества робастных настроек при наличии малой экспериментальной выборки состоит в том, чтобы найти в пространстве (k,T,τ) согласующуюся с экспериментами область $\boldsymbol{G}$, которая задана аналитически.

Рассмотрим, как и ранее, вектор $\boldsymbol{X}$ параметров модели объекта управления (в нашем случае $\boldsymbol{X} = [k,T,\tau]^T$). Предположим, что компоненты $\boldsymbol{X}$ подчинены нормальному распределению. Это допущение справедливо для ситуации, когда вариации параметров модели объекта являются суммой большого числа случайных факторов различной природы. Заметим, что интервальная неопределенность соответствует иной гипотезе, а именно: гипотезе равномерного распределения.

Из теории вероятностей известно, что совместная плотность вероятности системы случайных величин $(X_1,\dots,X_n)$ в случае нормального распределения имеет вид

$$f(x_1,\dots,x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\Delta}} exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} k_{ij}^{-1}(x_i - M_i)(x_j - M_j)\right\}, \qquad (15)$$

где $M_i -$ математическое ожидание X_i, $\Delta -$ определитель ковариационной матрицы $\boldsymbol{K} = [k_{ij}]$, k_{ij}^{-1} – элемент матрицы $\boldsymbol{K^{-1}} = [k_{ij}^{-1}]$ обратной по отношению к ковариационной матрице.

То же самое может быть более компактно записано в векторно-матричной форме, если обозначить $\boldsymbol{X} = (X_1,\dots,X_n)^T$. Тогда (15) запишется в виде

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\Delta}} exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{M_x})^T\boldsymbol{K^{-1}}(\boldsymbol{x}-\boldsymbol{M_x})\right\}.$$

Ясно, что

$$\max f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\Delta}}.$$

Теоретически случайные величины $X_1,\dots,X_n$ могут принимать значения в интервале $(-\infty,\infty)$, однако область пространства $(x_1,\dots,x_n)$, куда попадает большая часть реализаций случайного вектора $\boldsymbol{X}$, задается условием

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\Delta}} exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{M_x})^T\boldsymbol{K^{-1}}(\boldsymbol{x}-\boldsymbol{M_x})\right\} \geq c \qquad (16)$$

где c – достаточно малая константа, значение которой определяется допустимым риском выхода за границы (16).

Если руководствоваться тем, что должно быть

$$c \ll \max f(\boldsymbol{x}),$$

то можно принять

$$c = \alpha \cdot \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\Delta}},$$

где $0 < \alpha \leq 1$, например, $\alpha = 0.1$.

Применив сказанное к вектору $\boldsymbol{X} = [k,T,\tau]^T$ в совокупности с методом моментов [7], получим аналитическое описание области неопределенности G в виде

$$exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{M_x})^T\boldsymbol{K^{-1}}(\boldsymbol{x}-\boldsymbol{M_x})\right\} \geq \alpha.$$

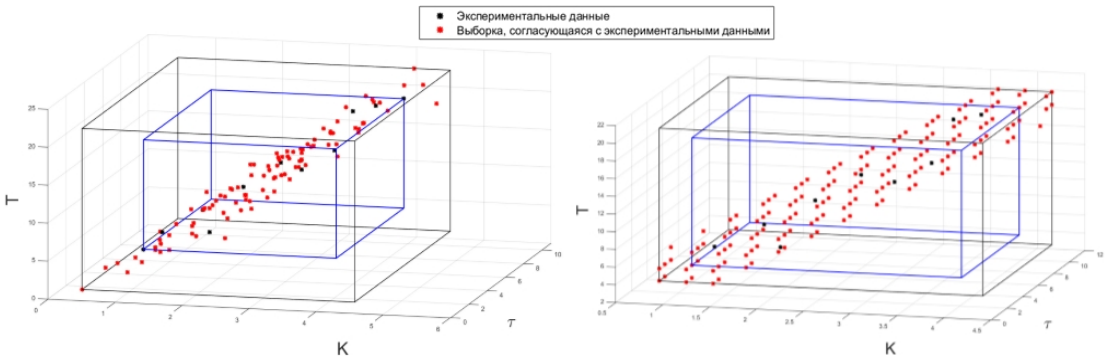
Далее задача расчета робастных настроек регулятора может быть решена методом аппроксимирующих параллелепипедов или способом точечной аппроксимации.

Анализ эффективности рассмотренных методов робастного управления. Для проверки эффективности предложенных методов рассматривался распространенный на практике вари-

ант, когда информация об области $\boldsymbol{G}$ задана набором экспериментально полученных точек в пространстве (k, T, τ). Исследование проводилось применительно к системе, управляемой ПИ-регулятором.

Рассматривалась выборка малого объема (11 значений), которая содержит данные экспериментов по идентификации модели интересующего объекта управления или сходных с ним объектов. На выход управляемого объекта подавалось ступенчатое возмущение, и моделировались переходные процессы в замкнутой системе.

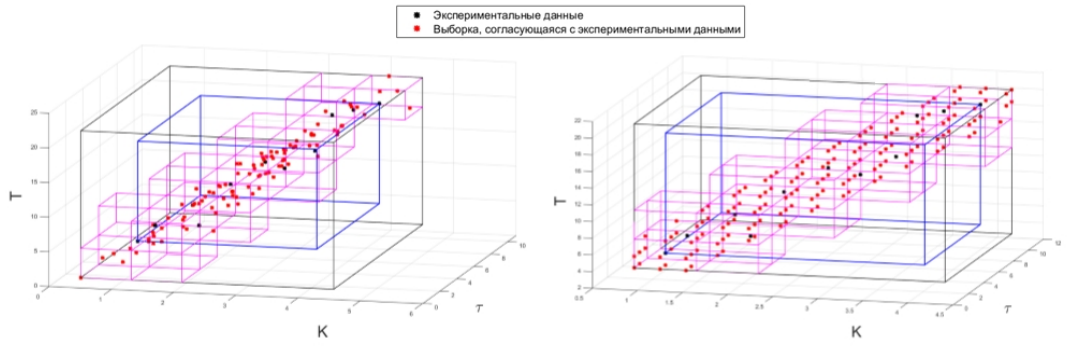
Методом Монте-Карло было произведено увеличение исходной малой выборки до 111 точек. Был применен также альтернативный способ восстановления вероятной зоны неопределенности с использованием многомерного нормального распределения (см. рис. 7).



a) b)

Рис. 7. a) Выборка, полученная с помощью метода Монте-Карло, b) Область, полученная с помощью применения многомерного нормального распределения

Далее для определения робастных настроек к полученным выборкам применялись МАП и МТА. (см. рис. 8)



a) b)

Рис. 8. Применение метода аппроксимирующих параллелепипедов и точечной аппроксимации к выборке, полученной с помощью a) метода Монте-Карло, b) многомерного нормального распределения

Таблица 1

Робастные настройки ПИ-регулятора и значения ИКК

	Монте-Карло		Исходная выборка		Многомерное нормальное распределение	
	МАП	МТА	МАП	МТА	МАП	МТА
Коэффициенты ПИ-регулятора	$k_{p_{МК_map}} = 0.0820$ $k_{i_{МК_map}} = 0.0150$	$k_{p_{МК_mta}} = 0.3350$ $k_{i_{МК_mta}} = 0.0171$	$k_{p_{ИВ_map}} = 0.3230$ $k_{i_{ИВ_map}} = 0.0194$	$k_{p_{ИВ_mta}} = 0.378$ $k_{i_{ИВ_mta}} = 0.019$	$k_{p_{МНР_map}} = 0.2360$ $k_{i_{МНР_map}} = 0.0168$	$k_{p_{МНР_mta}} = 0.3280$ $k_{i_{МНР_mta}} = 0.0168$
Среднее ИКК	0.215	0.1733	0.1697	0.1476	0.1837	0.174
Максимальное ИКК	0.249	0.210	0.199	0.184	0.220	0.212

Данные таблицы 1 подтверждают обсуждавшиеся ранее эффекты. Из таблицы видно, что робастные настройки, рассчитанные любым из рассмотренных методов, обеспечивают устойчивость замкнутой системы при любых имевших место или спрогнозированных с применением стохастических методов сочетаниях значений параметров динамической модели управляемого объекта.

Как и предполагалось, МАП приводит к более «осторожным» настройкам, чем МТА. Они, однако, обеспечивают большие гарантии получения допустимых показателей качества управления при любых возможных изменениях характеристик объекта. Ценой этих преимуществ может являться некоторое ухудшение средних и экстремальных показателей качества управления по сравнению со способом точечного задания или точечной аппроксимации множества $\boldsymbol{G}$.

Применение МАП и МТА к исходной малой выборке экспериментальных данных, казалось бы, приводит к лучшим результатам по ИКК, чем применение тех же процедур к расширенным выборкам, полученным методом Монте-Карло или путем точечной аппроксимации эллипсоида, рассчитанного с использованием трехмерного нормального распределения. Этот выигрыш, однако, может оказаться кажущимся – если предположить, что реальная область $\boldsymbol{G}$ совпадает с одной из двух сформированных виртуальных областей, то настройки регулятора, рассчитанные по малой выборке, не будут обладать свойством робастности. Дело в том, что благодаря построению виртуальных множеств удается «нащупать узкие места» возможного резкого ухудшения качества управления, т.е. те точки, значение ИКК в которых значительно превышает среднее.

Помимо разобранных выше методов, был рассмотрен и наиболее простой способ определения робастных настроек – метод окаймляющего параллелепипеда. Как было описано ранее, при реализации данного подхода робастной является точка, в которой достигаются максимальные коэффициент усиления k и запаздывание τ, а также минимальная инерционность T.

Видно, что такой подход может (в зависимости от расположения области неопределенности $\boldsymbol{G}$ внутри окаймляющего ее параллелепипеда) приводить к слишком слабым робастным настройкам. Это связано с тем, что метод окаймляющего параллелепипеда не учитывает зависимость между параметрами модели объекта управления. В рассматриваемом варианте $k_{p_{\text{МОП}}} = 0.0995,\ k_{i_{\text{МОП}}} = 0.0194,\ \text{ИКК}_{\text{сред}} = 0.2024,\ \text{ИКК}_{\text{макс}} = 0.2277.$

Заключение

Для распространенных в приложениях инерционных процессов с запаздыванием разработаны приближенные методы робастной настройки типовых регуляторов применительно к двум видам неопределенности. Для варианта интервальной неопределенности обоснованы простые расчетные соотношения. Для более сложного варианта неинтервальной неопределенности предложены и исследованы методами имитационного компьютерного моделирования метод окаймляющего параллелепипеда и более сложные, но и более эффективные методы аппроксимирующих параллелепипедов и точечной аппроксимации. Предложены и исследованы на численных примерах различные способы задания зоны неопределенности с учетом корреляции параметров модели объекта управления (способы, базирующиеся на методе Монте-Карло и на трехмерном нормальном распределении). Их положительная особенность состоит в том, что при наличии малых экспериментальных выборок они повышают гарантии сохранения допустимого качества управления за счет определенным образом организованного дополнения малого объема экспериментальных данных виртуальными данными, согласующимися с экспериментально выявленными корреляционными зависимостями.

ЛИТЕРАТУРА

1. Поляк Б.Т., Щербаков П.С. Робастная устойчивость и управление. // М: Наука, 2002, 303 с.
2. Гурецкий Х. Анализ и синтез систем управления сзапаздыванием. // М: Машиностроение, 1974,,328 с
3. PID Control for Multivariable Processes /Qing-Guo Wang [et al.] // Springer-Verlag, Berlin – Heidelberg, 2008, 261 p.
4. Яковис Л.М. Простые способы расчета типовых регуляторов для сложных объектов промышленной автоматизации // Автоматизация в промышленности № 6, 2007, с. 51 – 56.

5. Доронина Н.А., Яковис Л.М. Робастно-адаптивное управление инерционными объектами с запаздыванием // Неделя науки СПбПУ : материалы научной конференции с международным участием. Институт прикладной математики и механики – СПб.: Изд-во Политехн. ун-та, 2016, с. 62 – 64.
6. Кибардин С.А., Яковис Л.М. Робастное управление инерционными объектами с запаздыванием // XLI Неделя науки СПбГПУ: материалы научно-практической конференции с международным участием СПб.: Изд-во СПбГПУ, 2012, с. 71 – 72.
7. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений // М: Наука, 1969, 512 с..

L.M. Yakovis, N.A. Doronina (St. Petersburg Polytechnic University of Peter the Great, Saint Petersburg)

Accounting for the character of the parametric uncertainty of the dynamic model when calculating robust regulators for objects with self-leveling

The work deals with the problem setting the standard proportional-integral (PI) and a proportional-integral-differential (PID) controllers for sustainable dynamic objects with inertia and delay in an environment where the model parameters are not known precisely, and given only the region of possible values. Developed and tested with the help of simulation computer simulation efficient methods of robust adjustment of typical regulators. A comparative analysis of the effectiveness of the proposed algorithms is performed.

ОБОБЩЕНИЕ МЕТОДА АГРЕГИРОВАННЫХ РЕГУЛЯТОРОВ НА СТОХАСТИЧЕСКИЙ ДИСКРЕТНЫЙ ОБЪЕКТ И ЕГО ПРИМЕНЕНИЕ ДЛЯ РЕШЕНИЯ ПРИКЛАДНЫХ ЗАДАЧ

С. И. Колесникова[1], **С. А. Цветницкая**[2]
(*Национальный исследовательский Томский государственный университет, Томск, пр. Ленина, 36, skolesnikova@yandex.ru*)

П. А. Побегайло[3]
(*Институт машиноведения им. А.А. Благонравова РАН, Москва, пер. Малый Харитоньевский, 4, petrp214@yandex.ru*)

Аннотация

Дается постановка задачи синтеза робастного стохастического регулятора на основе принципов управления на многообразиях и новый алгоритм конструирования для стохастического дискретного объекта второго порядка.

Приведены иллюстративные примеры решения двух прикладных задач на базе сконструированного регулятора. В первом примере демонстрируется его применение для управления экономическим объектом, неустойчивым в разомкнутом состоянии.

Второй пример содержит теоретический результат в виде нового алгоритма синтеза системы управления векторным объектом типа «манипуляционная система», физическая модель которого есть рабочий орган одноковшового гидравлического экскаватора. Рассмотрена задача построения системы управления процессом стабилизации углового положения ковша.

Алгоритм конструирования системы управления и реализованный соответствующий алгоритм управления применительно к решению задачи стабилизация движения ковша экскаватора-робота по заданной поверхности могут быть полезны для любых манипуляционных систем.

Ключевые слова: робастный стохастический регулятор на многообразиях, управление стохастическим нелинейным объектом, управление манипуляционной системой.

Введение

Для решения задачи управления с параметрической неопределенностью в описании объекта в классической теории адаптивного управления постулируется существование целевого «идеального» набора параметров (устранимая неопределенность). Предметом теории робастного управления являются возмущения (в объекте, в регуляторе, неизвестные детерминированные, стохастические и параметрические шумы), характеризующие неточность описания, нелинейности (неустранимая неопределенность) [1-9]. Так, в объекте с описанием $Y_{i+1} = F_i + u_i + \xi_{i+1} + c\xi_i$, где $\{\xi_i\}_{i\geq 1}$ независимые одинаково распределенные случайные величины с характеристикам $M\xi_i = 0, D\xi_i = \sigma^2,\ 0 < c < 1,\ i \geq 1,\ i \in \mathrm{N}$, $F_i = F(Y_i)$, u_i - нелинейная функция и управление, соответственно, дисперсия выходного сигнала $D(Y_{i+1}) = D(F_i + u_i + c\xi_i) + \sigma^2 \geq \sigma^2$, не может быть меньше дисперсии присутствующего в описании объекта шума при любом управлении. В этом случае «разумно» потребовать нахождения такого управления, при котором дисперсия выходной переменной будет минимальной [2].

Доклад посвящен вопросу распространения классического метода аналитического конструирования агрегированных регуляторов (АКАР) [10], разработанного ранее только для детерминированного случая, на нелинейные объекты со стохастической неопределенностью с целью получения робастной системы управления.

В первой части доклада излагается алгоритм синтеза робастного регулятора на основе стохастического метода аналитического конструирования агрегированных регуляторов для объекта второго порядка; во второй части приводятся примеры решения двух разноплановых прикладных задач на единой основе: управления экономической моделью рынка и рабочим оборудованием экскаватора.

[1] Доктор технических наук, профессор.
[2] Кандидат технических наук, доцент.
[3] Кандидат технических наук, старший научный сотрудник.

Постановка задачи синтеза робастного дискретного регулятора на основе многообразий

Пусть объект задан системой нелинейных стохастических разностных уравнений:

$$\begin{aligned} X_1[k+1] &= F_1[k] + \xi_1[k+1] + c_1\xi_1[k], \\ X_2[k+1] &= F_2[k] + u[k] + \xi_2[k+1] + c_2\xi_2[k], \end{aligned} \tag{1}$$

где $F_i[k] := F_i\left(X_1[k], X_2[k], \theta\right)$, $i = 1,2$ - нелинейные ограниченные функции с вектором параметров θ; вектор $X[t_k] := X[k] = \left(X_1[k], X_2[k]\right), t_k = k\Delta, k \in \{0,1,...\}, \Delta > 0$ измеряется по закону $Y[k] = X[k], k \in \{0,1,...\}$ в моменты $t_k, k \in \{0,1,...\}$; $u[k]$ - искомое управление на полуинтервале $[t_k, t_{k+1})$; управление производится на промежутке $[t_0, \infty)$; $\{\xi_i[k]\}_{k\geq 0}$, $i = 1,2$ – последовательности независимых одинаково распределенных случайных величин со средним 0 и дисперсиями σ_i^2, $i = 1,2$, соответственно; коэффициенты $0 < c_i < 1$, $i = 1,2$ характеризуют скорость «забывания предыстории».

Ставится задача синтеза робастного управления, выводящего объект (1) в целевое множество состояний, описанное в виде среднего значения целевого многообразия $\mathbf{E}\{\psi[k]\} = 0, \psi[k] = \psi\left(X_1[k], X_2[k]\right) = 0, k \in \{0,1,...\}$, где $\psi(\cdot)$ - известная функция состояния, называемая макропеременной.

При этом требуется, чтобы $\mathbf{E}\{\psi[k+1] + \omega\psi[k]\} = 0$, а дисперсии $\mathbf{D}\{\psi[k+1] + \omega\psi[k]\}$, $\mathbf{D}\{X_2[k+1]\}$ случайных величин $\psi[k+1] + \omega\psi[k]$ и выходной макропеременной $\psi\left(X_1[k], X_2[k]\right)$ были равны наименьшим значениям из возможных $\forall k \geq 0$ при заданных начальных условиях $\left(X_1[0], X_2[0]\right)$.

Для читабельности и наглядности сути алгоритма синтеза робастного стохастического регулятора рассмотрим две задачи управления с описанием целей управления, соответственно:

$$\psi[k] = X_1[k] - \rho X_2[k] = 0,\ k \to \infty,\ \rho \neq 0, \tag{2}$$

$$\psi[k] = X_1[k] - X_1^*[k] = 0,\ k \to \infty. \tag{3}$$

Задачей управления 1 назовем совокупности описаний (1) и (2); *задачей управления 2* - описания (1) при условии $\xi_1[k] = 0, k \in \{0,1,...\}$ и цели (3). Здесь $X_1^*[k]$ - известная целевая траектория.

Решение задачи синтеза робастного дискретного регулятора

Решение задачи 1 реализуется двумя шагами. На *первом* получаем вид базового АКАР-управления $\tilde{u}[k]$ при условии, что величины $\{\xi_i[k]\}_{k\geq 0}$, $i = 1,2$ известны. Из техники классического метода АКАР следует, что

$$\tilde{u}[k] = -F_2[k] - \xi_2[k+1] - c_2\xi_2[k] + \rho^{-1}\left(F_1[k] + \xi_1[k+1] + c_1\xi_1[k] + \omega\psi[k]\right). \tag{4}$$

Утверждение 1. Управление (4) для задачи 1 при известных $\{\xi_i[k]\}_{k\geq 0}$, $i = 1,2$ обеспечивает асимптотически устойчивое поведение системе (1), (2), (4) и доставляет решение вариационной задаче: $J = \sum_{k=0}^{\infty}\left(\alpha^2\psi^2[k] + \left(\Delta\psi[k]\right)^2\right) \to \min, \psi = 0$. При этом уравнение $\psi[k+1] + \omega\psi[k] = 0,\ |\omega| < 1,\ k \geq 0$ есть уравнение Эйлера-Лагранжа для J_D, а скаляры α и ω связаны соотношением: $\omega = 0,5\left(1 + \alpha^2 - \sqrt{\left(1+\alpha^2\right) - 4}\right), |\omega| < 1$.

Доказательство утверждения 1 непосредственно следует из дискретного аналога АКАР (см., например, [10]).

Второй шаг опирается на принцип учета апостериорной информации в виде условного математического ожидания на основе (4):

$$u[k]=\mathbf{E}\left\{\tilde{u}[k]\,\middle|\,\xi^k\right\}=\rho^{-1}F_1[k]-F_2[k]+\rho^{-1}\lambda\psi[k]-c_2\xi_2[k]+\rho^{-1}c_1\xi_1[k], \tag{5}$$

где $\xi^k=\left(\xi[k],\xi[k-1]...\xi[0]\right),\xi[k]=\left(\xi_1[k],\xi_2[k]\right),\ k\in\{0,1,...\}$.

Утверждение 2. В условиях закона управления (5) (в смысле **P**-п.н.) для выходной переменной $X_2[k]$ и макропеременной (2) имеем соотношения:

$$\psi[k+1]+\omega\psi[k]=\xi_1[k+1]-\rho\xi_2[k+1]. \tag{6}$$

$$X_2[k+1]=\rho^{-1}F_1[k]+\rho^{-1}\omega\psi[k]+\rho^{-1}c_1\xi_1[k]+\xi_2[k+1]. \tag{7}$$

Для получения итогового вида регулятора для задачи 1 следует из исключить выражение $-c_2\xi_2[k]+\rho^{-1}c_1\xi_1[k]$ в (5) на основе уравнений (6), (7). Положив для упрощения решения этой задачи $c_1=c_2=c$ с учетом (6), (7) получим:

$$u[k]=\rho^{-1}F_1[k]-F_2[k]+\left(\rho^{-1}\omega+c\rho\right)\psi[k]+c\rho\lambda\psi[k-1]. \tag{8}$$

Итоговая система управления в пространстве состояний объекта (1) для задачи 2 имеет вид совокупности уравнений (1), (2), (8):

$$\begin{aligned}
&X_1[k+1]=F_1[k]+\xi_1[k+1]+c_1\xi_1[k],\\
&X_2[k+1]=F_2[k]+u[k]+\xi_2[k+1]+c_2\xi_2[k],\\
&u[k]=\rho^{-1}F_1[k]-F_2[k]+\left(\rho^{-1}\omega+c\rho\right)\psi[k]+c\rho\omega\psi[k-1],\\
&\psi[k]=X_1[k]-\rho X_2[k],\ \rho\neq 0,\ |\omega|<1,\ k\geq 0.
\end{aligned} \tag{9}$$

Утверждение 3. Управление (8) для задачи 1 обеспечивает асимптотически устойчивое в среднем поведение системы (9) в окрестности многообразия $\psi[k]=X_1[k]-\rho X_2[k]=0$ и обладает свойствами:

1) $\mathbf{E}\left\{\psi[k+1]+\omega\psi[k]\right\}=0$;
2) дисперсии случайных величин $\psi[k+1]+\omega\psi[k]$ и выходной величины $\psi[k]=X_1[k]-\rho X_2[k]$ под воздействием управления (8) равны наименьшим из возможных значений $\forall k\geq 0$ и подчиняются соотношениям:

$$\begin{aligned}
&\mathbf{D}\left\{\psi[k+1]+\omega\psi[k]\right\}=\sigma_1^2+\rho^2\sigma_2^2,\\
&\mathbf{D}\left\{\psi[k+1]\right\}=\sigma_1^2+\rho^2\sigma_2^2+\omega^2\mathbf{D}\left\{\psi[k]\right\},\ |\omega|<1,k\geq 0.
\end{aligned} \tag{10}$$

Доказательство утверждения 3 осуществляется непосредственной проверкой.

Решение задачи 2. На первом шаге согласно классическому методу АКАР при условии, что правая часть описания (1) известна и с учетом предположения $\xi_1[k]=0,\ \xi[k]=\xi_2[k],\ k\in\{0,1,...\}$ и вида цели (3), получаем описание структуры системы управления в виде:

$$\begin{aligned}
&X_1[k+1]=F_1[k],\ F_1[k]=F_1\left(X_1[k],X_2[k]\right),\\
&X_2[k+1]=F_2[k]+u[k]+\xi[k+1]+c\xi[k],\ F_2[k]=F_2\left(X_1[k],X_2[k]\right),\\
&\psi[k]=X_1[k]-X_1^*[k],\ \psi^{(1)}[k]=X_2[k]-\varphi\left(X_1[k]\right),\\
&\tilde{u}[k]=-F_2[k]-\xi[k+1]-c\xi[k]+\varphi\left(F_1[k]\right)-\lambda_1\psi^{(1)}[k],\\
&F_1\left(X_1[k],\varphi\left(X_1[k]\right)\right)-X_1^*[k+1]+\lambda_2\psi[k]=0,\ k\in\{0,1,...\}.
\end{aligned} \tag{11}$$

В описании (11) функция $\varphi\left(X_1[k]\right)$, согласно АКАР интерпретируемая как «внутреннее управление», находится из последнего уравнения формул (11).

На втором шаге управление ищется в виде условного математического ожидания $u[k]=\mathbf{E}\{\tilde{u}[k]\,|\,\xi^k\}$, $k\in\{0,1,...\}$, явное определение которого в общем виде весьма затруднительно, поэтому для наглядности изложим этот шаг в процессе решения прикладной задачи 2.

Решение прикладных задач синтеза робастного регулятора для дискретной модели

Прикладная задача 1. Рассмотрим экономический объект, построенный на основе модели Фейгенбаума (которая среди предельных состояний имеет хаотические) и исследованную на устойчивость в [10] (стр. 447) в условиях отсутствия управления:

$$\begin{aligned} &X_1[k+1]=F_1[k],\ F_1[k]=X_1[k]\left(\alpha C_0-\mu\beta_1 X_1[k]X_2[k]\right),\\ &X_2[k+1]=F_2[k],\ F_2[k]=A\left(\alpha C_0-\mu\beta_2 X_1[k]X_2[k]\right),\ k\in\{0,1,...\}. \end{aligned} \tag{12}$$

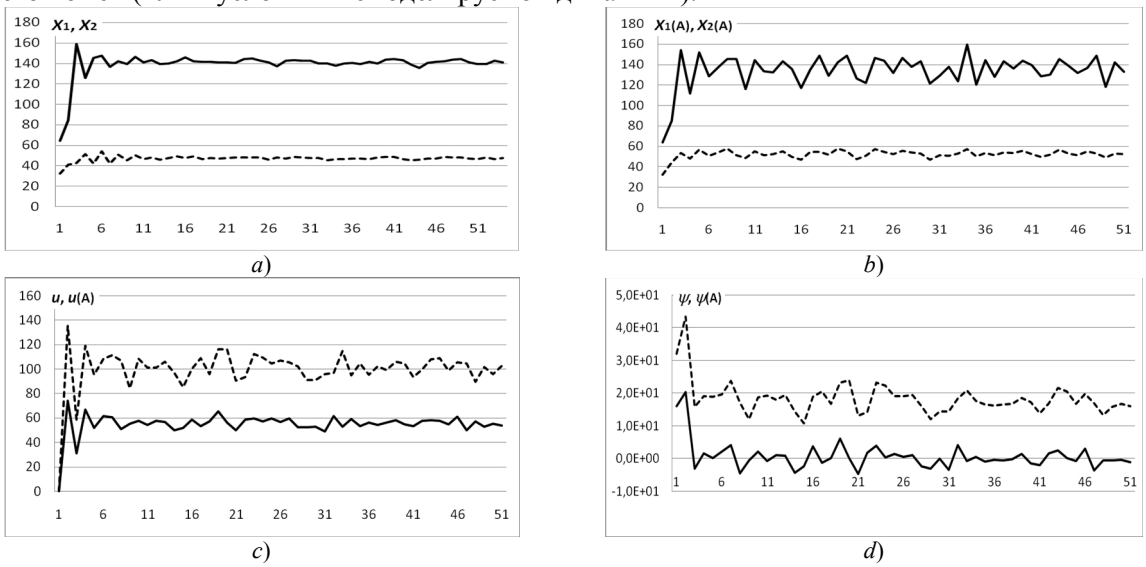
Содержательный смысл этой модели может отражать динамику достаточно многих показателей [11-13] как технических, так и социальных, экономических и биологических. Так, здесь переменные $X_1[k],\ X_2[k]$ интерпретируются как величины, пропорциональные объемам продукции, поставляемые на рынок со стороны субъектов $X_1,\ X_2$, соответственно; параметры α, C_0, μ - коэффициенты пропорциональности с содержательным экономическим смыслом; параметры β_1,β_2 означают цены, устанавливаемые производителями однотипной продукции $X_1,\ X_2$, соответственно.

Отметим, что ранее на основе описания (12) был успешно построен расширенный вариант объекта *детерминированного управления* 4-го порядка [14, 15], где управляемыми переменными полагались цены поставляемой на рынок продукции, что физически реализуемо.

Здесь для иллюстрации решения задачи управления 1 рассмотрим один из вариантов аддитивно входящего управления для зашумленного в смысле описания (1) объекта:

$$\begin{aligned} &X_1[k+1]=X_1[k]\left(\alpha C_0-\mu\beta_1 X_1[k]X_2[k]\right)+\xi_1[k+1]+c_1\xi_1[k],\\ &X_2[k+1]=A[k]\left(\alpha C_0-\mu\beta_2 X_1[k]X_2[k]\right)+u[k]+\xi_2[k+1]+c_2\xi_2[k],\ k\in\{0,1,...\}. \end{aligned} \tag{12}$$

На основе решения задачи управления 1 с целью управления (2), получим систему *стохастического управления* в виде (9) с учетом конкретных описаний $F_1[k],\ F_2[k]$, поведение которой иллюстрируется на рисунках 1, *a*)-*d*) из которых следует явная предпочтительность использования стохастического регулятора перед детерминированным в условиях нерасчетных для него помех (или в условиях немоделируемой динамики).



a) *b*) *c*) *d*)

Рис. 1. Траектории *a*) переменных объекта стохастического управления $X_2[k]$ - сплошное начертание, $X_1[k]$ – пунктир; *b*) переменной объекта детерминированного АКАР-управления в условиях помех; *c*) управляемых переменных: сплошное начертание для стохастического управления $u[k]$, пунктир для детерминированного АКАР-управления $u(A)[k]$; *d*) макропеременных $\psi[k],\psi(A)[k]$, сплошное начертание и пунктир, соответственно

Прикладная задача 2. Рассмотрим технический объект «рабочее оборудование одноковшового гидравлического экскаватора», модель которого исторически построена на непрерывной основе (например, [16]), а управление находилось в скользящем режиме [17]:

$$m\ddot{b}(t) = q\left(b(t), \dot{b}(t)\right) + u,\ t \geq 0, \tag{13}$$

где $b \in R^3$ - вектор состояния механизма с координатами $b(t) = \left(b_1(t), b_2(t), b_3(t)\right)^T$, где $b_1 = \alpha(t), b_2 = \beta(t), b_3 = \gamma(t)$ - шарнирные (межзвенные) углы (рис. 2); $q = q\left(b, \dot{b}\right) \in R^3$ - так называемые обобщенные силы, действующие на объект управления; $m = \left\| m_{ij} \right\|_{3\times 3}$ - матрица кинетической энергии; $u \in R^3$, $u = u\left(b(t), \dot{b}(t)\right)$ - вектор управления (шарнирных моментов, $u_i = M_i$, $i = 1,2,3$).

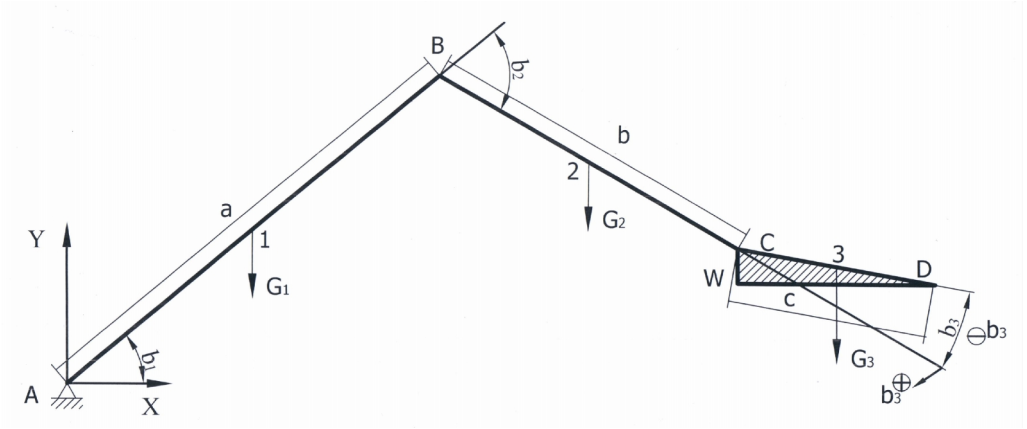


Рис. 2. Упрощенная кинематическая схема рабочего оборудования одноковшового гидравлического экскаватора прямого копания с независимыми угловыми перемещениями стрелы, рукояти и ковша

На рисунке использованы обозначения: XAY - декартова система координат с правой ориентацией; в точках A, B, C находятся вращательные шарниры, соединяющие элементы рабочего оборудования между собой и с базовой машиной (точка A); точка D - концевая точка зубьев ковша рабочего оборудования; в точке W находится место крепления гидроцилиндров механизма привода ковша к ковшу (здесь играет иллюстративно-вспомогательную роль); AB, BC, CD - звенья рабочего оборудования экскаватора и их кинематические длины a, b, c, соответственно; точки 1, 2 и 3 обозначают центры тяжести соответствующих звеньев рабочего оборудования экскаватора (здесь они лежат в срединной точке звеньев); G_1, G_2, G_3 - веса каждого из звеньев рабочего оборудования; b_1, b_2, b_3 - шарнирные углы.

Ставится классическая задача стабилизации: требуется сконструировать управление, выводящее траектории объекта $b_i(t)$ на заданные $b_i^*(t)$, $i = 1,2,3$.

Поскольку «тяготение к дискретности работы [сложно организованных систем] вызвано достаточно разъясненными в настоящее время причинами» [18], то для применения техники конструирования робастного стохастического регулятора построим дискретную модель на основе описания (13), что не противоречит практике, поскольку определение значений управляющий воздействий в любом вычислительном пакете осуществляется по дискретной схеме («весьма вероятно, что в очень многих случаях разумно изучение реальных явлений вести, избегая промежуточный этап их стилизации в духе представлений математики бесконечного и непрерывного, переходя прямо к дискретным моделям» (по А.Н. Колмогорову)).

Введем сначала обозначения: $x_1(t) = b(t)$, $x_2(t) = \dot{b}(t)$, где $x_1\left(t\right) = b\left(t\right)$, $x_2\left(t\right) = \dot{b}(t)$, $x_1 = \left(x_{11}, x_{12}, x_{13}\right)^T$, $x_2 = \left(x_{21}, x_{22}, x_{23}\right)^T$. Исходную модель (в матричной форме) преобразуем к виду, удобному для применения базового алгоритма аналитического конструирования регулятора:

$$\begin{aligned} x_1[k+1] &= x_1[k] + \tau x_2[k], \\ x_2[k+1] &= x_2[k] + \tau aq\left(x_1[k], x_2[k]\right) + \tau au[k] + \xi[k+1] + c\xi[k],\ k \in \{0,1,...\} \end{aligned} \tag{14}$$

где $a = m^{-1}$, x_i, q, $u \in R^3$, $i = 1,2$, $m = \|m_{ij}\|_{3\times 3}$; случайные величины при фиксированном k $\{\xi[k]\}_{k\geq 0}$ и постоянная c обладают теми же свойствами, что и при описании объекта (1); τ - коэффициент пропорциональности.

Замечание 1. Физическая интерпретация случайных функций в правой части системы (14) характеризуют погрешности работы оборудования, измерительных приборов и погрешности управления.

Осуществляем сначала (по аналогии с первой частью текста доклада) детерминированный АКАР-синтез для объекта (14) с целью управления:

$$\psi\left(x_1[k]\right) = x_1[k] - b^*[k] = 0, \psi \in R^3, \psi = \left(\psi_1, \psi_2, \psi_3\right)^T,\ k \geq 0 \tag{15}$$

и функционалом качества синтезируемой системы управления:

$$\Phi = \sum_{k=0}^{\infty}\sum_{i=1}^{3}\left(\alpha_i^2\psi_i^2[k] + \left(\Delta\psi_i[k]\right)^2\right) \to \min. \tag{16}$$

Далее вариационную задачу вида (15), (16) будем для удобства обозначать парой символов $(\Phi,\ \psi)$.

Решение первого шага будет более громоздким в силу того, что целевые переменные не являются управляемыми, и представляет самостоятельный теоретический и практический интерес.

1. Вывод структуры системы управления. С этой целью вводится сначала вспомогательная макропеременная:

$$\psi^{(\mathrm{I})}[k] = x_2[k] - \varphi\left(x_1[k]\right),\ k \in \{0,1,...\}, \tag{17}$$

называемая, согласно классическому АКАР, «внутренним» управлением. Здесь $\psi^{(\mathrm{I})} = \left(\psi_1^{(\mathrm{I})}, \psi_2^{(\mathrm{I})}, \psi_3^{(\mathrm{I})}\right)$; целью управления является множество $\left\{x : \psi^{(\mathrm{I})}(x) = 0\right\}$.

Ставится первая вариационная задача $\left(\Phi_1,\ \psi^{(1)}\right)$, сопровождающая синтез системы управления, где $\Phi_1 = \sum_{k=0}^{\infty}\sum_{l=1}^{3}\left(\left(\alpha_l^{(1)}\right)^2\left(\psi_l^{(1)}[k]\right)^2 + \left(\Delta\psi_l^{(1)}[k]\right)^2\right) \to \min$ и ограничения $\psi^{(\mathrm{I})} = 0$.

Управления $\tilde{u}_i$, $i = 1,2,3$ (при известных $\{\xi[k]\}_{k\geq 0}$) ищутся из матричного уравнения $\psi^{(\mathrm{I})}[k+1] + \omega_1\psi^{(\mathrm{I})}[k] = 0$, $\psi^{(\mathrm{I})}[k] \in R^3$, $0<\omega_1<1$, $k \in \{0,1,...\}$, на решениях которого, в свою очередь, достигается безусловный минимум функционала Φ_1. В итоге получаем структуру управления с точностью до (пока неизвестной) функции $\varphi(x_1)$ в виде:

$$\begin{aligned} \tau a\tilde{u}[k] &= -F_2[k] - \xi_2[k+1] - c_2\xi_2[k] + \varphi\left(F_1[k]\right) - \omega_1\psi^{(1)}[k], \\ F_2[k] &= x_2[k] + \tau aq\left(x_1[k], x_2[k]\right),\ k \in \{0,1,...\}. \end{aligned} \tag{18}$$

Содержательно множество $\left\{x : \psi^{(\mathrm{I})}(x) = 0\right\}$ несет смысл первого множества состояний с аттрактивным (притягивающим) свойством для решений системы уравнений (14), в том смысле, что из многих вариантов возможного поведения нелинейного объекта (14) выбирается то, которое обусловлено свойством $\psi^{(\mathrm{I})}(x) = 0$.

Далее для получения вида $\varphi(x_1)$ осуществляется редукция (декомпозиция) объекта (14) на многообразии $\psi^{(\mathrm{I})} = 0$, из выполнения которого следует $x_2 = \varphi(x_1)$ и, соответственно, редуцированная система примет вид:

$$\hat{x}_1[k+1] = \hat{x}_1[k] + \tau\varphi\left(\hat{x}_1[k]\right),\ k \in \{0,1,...\}. \tag{19}$$

Замечание 2. Переменные в описании (19) несут смысл поведения (14) на многообразии $\psi^{(\mathrm{I})}=0$, и, вообще говоря, уже другие, поэтому здесь используется обозначение $\hat{x}_1 := x_1$.

Для решения задачи определения управления $\varphi(x_1)$ вводится макропеременная следующего типа:

$$\psi^{(\mathrm{II})}[k]=\psi^*\left(\hat{x}[k]\right)=\hat{x}_1[k]-b^*[k],\ k\in\{0,1,...\}. \tag{20}$$

и ставится вторая вариационная задача $\left(\Phi_2,\ \psi^{(\mathrm{II})}\right)$, сопровождающая синтез системы управления, в виде критерия $\Phi_2=\sum_{k=0}^{\infty}\sum_{l=1}^{3}\left(\left(\alpha_l^{(\mathrm{II})}\right)^2\left(\psi_l^{(\mathrm{II})}[k]\right)^2+\left(\Delta\psi_l^{(\mathrm{II})}[k]\right)^2\right)\to\min$ и ограничений вида $\psi^{(\mathrm{II})}=0,\ \psi^{(\mathrm{II})}=\left(\psi_1^{(\mathrm{II})},\psi_2^{(\mathrm{II})},\psi_3^{(\mathrm{II})}\right)$.

Внутреннее управление $\varphi(\hat{x}_1)$ ищется из решениях матричного уравнения $\psi^{(\mathrm{II})}[k+1]+\omega_2\psi^{(\mathrm{II})}[k]=0,\ \psi^{(\mathrm{II})}[k]\in R^3,\ 0<\omega_2<1$, на которых, в свою очередь, достигается безусловный минимум функционала Φ_2. Из этого уравнения с учетом уже описаний (19), (20) получаем:

$$\psi^{(\mathrm{II})}[k+1]+\omega_2\psi^{(\mathrm{II})}[k]=\hat{x}_1[k+1]-b^*[k+1]+\omega_2\left(\hat{x}_1[k]-b^*[k]\right)=$$
$$=\hat{x}_1[k]+\tau\varphi\left(\hat{x}_1[k]\right)-b^*[k+1]+\omega_2\left(\hat{x}_1[k]-b^*[k]\right)=0,\ k\in\{0,1,...\}.$$

Из последнего уравнения управление $\varphi(x_1)$ будет иметь вид:

$$\varphi\left(\hat{x}_1[k]\right)=-\tau^{-1}\left(\hat{x}_1[k]-b^*[k+1]+\omega_2\left(\hat{x}_1[k]-b^*[k]\right)\right),\ k\in\{0,1,...\}. \tag{21}$$

Таким образом, итоговая система управления для достижения цели управления (15) или $b_i[k]-b_i^*[k]=0,\ i=1,2,3,\ k\to\infty$ объектом (14) *при известных шумах*, имеет вид совокупности уравнений (14), (15), (18), (21):

$$\begin{aligned}
&x_1[k+1]=x_1[k]+\tau x_2[k],\\
&x_2[k+1]=x_2[k]+\tau aq\left(x_1[k],x_2[k]\right)+\tau a\tilde{u}[k]+\xi[k+1]+c\xi[k],\\
&\psi\left(x_1[k]\right)=x_1[k]-b^*[k],\ \psi^{(\mathrm{I})}[k]=x_2[k]-\varphi\left(x_1[k]\right),\\
&\tau a\tilde{u}[k]=-F_2[k]-\xi[k+1]-c\xi[k]+\varphi\left(F_1[k]\right)-\omega_1\psi^{(1)}[k],\\
&F_2[k]=x_2[k]+\tau aq\left(x_1[k],x_2[k]\right),\\
&\varphi\left(x_1[k]\right)=-\tau^{-1}\left(x_1[k]-b^*[k+1]+\omega_2\left(x_1[k]-b^*[k]\right)\right),\ k\ge 0.
\end{aligned} \tag{22}$$

Утверждение 4. Система управления (22) в детерминированных условиях $\left(\xi[k]=0,k\ge 0\right)$ гарантирует достижение цели управления в виде: $\psi\left(x_1[k]\right)=x_1[k]-b^*[k]=0,\psi\in R^3,\ k\ge 0$ за конечное время и глобальный минимум функционалу качества управления вида $\Phi=\sum_{k=0}^{\infty}\sum_{l=1}^{3}\left(\left(\alpha_l\right)^2\left(\psi_l[k]\right)^2+\left(\Delta\psi_l[k]\right)^2\right)\to\min$, где $b^*[k],\ k\ge 0$ - заданная динамика изменения шарнирных углов, а параметр $\alpha\in R^3$ - коэффициент пропорциональности длительности переходного процесса.

На рис. 3 приведены результаты моделирования при известной правой части, свидетельствующие о корректной работе системы управления (22). Траектории переменных объекта $x_1[k]\in R^3,\ k\ge 0$ детерминированного управления (18) начинаются в начальном положении $x_1(0)=(0{,}997;\ 2{,}129;\ 0{,}14)$, что соответствует углам $b(0)=(56;\ 122;\ 8)$ (град.). Целевые значения $x_1^*(0)=(-0{,}052;\ -1{,}125;\ 0)$ достигнуты.

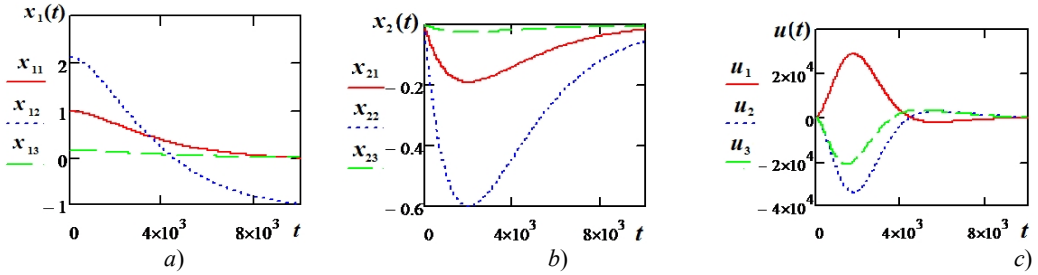


Рис. 3. *a*), *b*) - переходные процессы для фазовых координат системы (22) при условии полного описания, их производных, соответственно; *c*) - поведение переменной управления

2. Вывод стохастического регулятора для объекта (14). Управление для объекта в условиях неизвестных (ограниченных) шумов, как и в первой части доклада, ищется в виде условного математического ожидания $u[k]=\mathbf{E}\{\tilde{u}[k]\,|\,\xi^k\}$, $k\in\{0,1,...\}$ с учетом (17):

$$\begin{aligned}\tau a u[k]&=\mathbf{E}\{\tau a\tilde{u}[k]\,|\,\xi^k\}=\mathbf{E}\{-F_2[k]-\xi[k+1]-c\xi[k]+\varphi(F_1[k])-\omega_1\psi^{(1)}[k]\,|\,\xi^k\}=\\&=-F_2[k]-c\xi[k]-\omega_1 x_2[k]+E[k],\;\; E[k]=\mathbf{E}\{\varphi(F_1[k])+\omega_1\varphi(x_1[k])\,|\,\xi^k\}.\end{aligned}\quad(23)$$

С учетом вида входящего в правую часть условного математического ожидания на основе (21)

$$E[k]=-\tau^{-1}\left[(1+\omega_2)(x_1[k]+\tau x_2[k])+b^*[k+2]+\omega_2 b^*[k+1]\right]+\omega_1\varphi(X_1[k]).\quad(24)$$

Исключаем шум из второго уравнения (14) после подстановки в него полученного выражения для управления (23):

$$\xi[k+1]=x_2[k+1]+\omega_1 x_2[k]-E[k].\quad(25)$$

Подставляем $\xi[k]=x_2[k]+\omega_1 x_2[k-1]-E[k-1]$ из (25) в (23), получаем

$$\tau a u[k]=-F_2[k]-(c+\omega_1)x_2[k]-c\omega_1 x_2[k-1]+E[k]+cE[k-1].\quad(26)$$

Итоговое описание стохастической системы управления составляет совокупность уравнений (22), (24), (26).

В докладе численно исследуются свойства стохастического регулятора (26).

Утверждение 5. Управление (26) для прикладной задачи 2 обеспечивает асимптотически устойчивое в среднем поведение системы (14) в окрестности многообразия $\psi(x_1[k])=x_1[k]-b^*[k]=0$ и обладает свойствами:

1) $\mathbf{E}\{\psi[k+1]+\omega\psi[k]\}=0$;
2) дисперсии случайных величин $\psi[k]$ и выходной величины $X_1[k]$ под воздействием управления (26) равны наименьшим из возможных значений.

Заключение

В докладе изложен новый метод построения робастного стохастического регулятора, обладающего весьма привлекательными следующими свойствами:

1) простота и прозрачность техники синтеза структуры управления, обусловленная свойствами классического метода аналитического конструирования агрегированных регуляторов;
2) устойчивое в среднем достижение окрестности заданного множества состояний за конечное время;
3) обеспечение минимальной дисперсии выходной переменной и минимальной дисперсии левой части уравнения Эйлера-Лагранжа, на решениях которого строится структура системы стохастического управления;
4) явное предпочтение перед детерминированным методом аналитического конструирования агрегированных регуляторов в случае присутствия немоделируемой стохастической динамики в объекте управления.

Приведенные из разных прикладных областей примеры свидетельствуют о возможности разносторонней применимости обсуждаемой в докладе техники синтеза робастного стохастического регулятора [19], в частности, в системах поддержки принятия решений в задачах управления обширным спектром сложных объектов.

Работа проводилась при поддержке гранта РФФИ № 17-08-00920

ЛИТЕРАТУРА

1. Красовский А. А. Математическая и прикладная теория. Избранные труды. М.: Наука, 2002. 362 с.
2. Astroem K.J., Wittenmark B. Adaptive control. Нью-Йорк: Dover Publications, 2008. 590 p.
3. Freeman R.A., Kokotovic P.V. Robust nonlinear control design. Boston: Birkhäuser, 1996. 258 p.
4. Isidori A. Robust Feedback Design for Nonlinear Systems: a Survey // Turkish Journal of Electrical Engineering and Computer Sciences. 2010. Vol.18. P. 693-714.
5. Тюкин И.Ю., Терехов В.А. Адаптация в нелинейных динамических системах. СПб.: ЛКИ, 2008. 384 с.
6. Khalil H.K. Nonlinear systems. Prentice Hall, 2002. 750 p.
7. A. Astolfi, D. Karagiannis, R. Ortega. Nonlinear and Adaptive Control with Applications. Springer, 2008. 290 p.
8. Sepulchre R., Jankovic M., Kokotovic P.V. Constructive nonlinear control. Springer Science & Business Media, 2012. 313 p.
9. Krstić M., Kanellakopoulos I., Kokotović P. V. Nonlinear and adaptive control design. New York: John Wiley & Sons, 1995. 563 p.
10. Колесников А. А. Синергетика и проблемы теории управления: сборник научных трудов / Под ред. А.А. Колесникова. М.: ФИЗМАТЛИТ, 2004. 504 с.
11. Wei-Bin Zhang. Synergetic Economics: Time and Change in Nonlinear Economics // Springer Science & Business Media. 2013. p. 246.
12. Малинецкий Г.Г. Математические основы синергетики: Хаос, структуры, вычислительный эксперимент. М.: Либроком, 2009. 312 с.
13. Prokhorov A. Nonlinear dynamics and chaos theory in economics: a historical perspective // Quantile. 2008. No.4. P. 79–92.
14. Kolesnikova S., Mylnikova E. Application of nonlinear adaptation method for discrete economic objects // Proc. International Conference on Applied Mechanics and Mechanical Automation. DEStech Publications, Inc. P. 349-355. URL: DOI 10.12783/dtetr/amma2017/13385.
15. Колесникова С.И.. Система множественного управления для нелинейного объекта с неполным описанием // Труды СПИИРАН. 2017. № 6 (55). С. 114-133.
16. Матюхин В.И. Стабилизация движений манипулятора вдоль заданной поверхности // Автоматика и телемеханика. 2011. № 4. С. 71 – 85.
17. V.I. Utkin. Sliding modes in control and optimization. Springer, 2012. 286 p.
18. Колмогоров А.Н. Комбинаторные основания теории информации // Успехи математических наук. 1983. Т. 38. № 4. С. 27–36.
19. Kolesnikova S.I. A multiple-control system for nonlinear discrete object under uncertainty. Optimization Methods and Software. 2018. URL: https://doi.org/10.1080/10556788.2018.1472258.

S.I. Kolesnikova, S.A. Tsvetnitskaya (National Research Tomsk State University, Institute of Applied Mathematics and Computer Science, Tomsk), **P.A. Pobegailo** (A.A. Blagonravov Mechanical Engineering Research Institute of RAS, Moscow)

Expansion of the method of aggregated regulators on a stochastic discrete object and its application for solving applied problems

The statement of the problem of synthesizing a robust stochastic regulator based on the principles of control on manifolds and a new design algorithm for a second-order stochastic discrete object are given. Illustrative examples of solving two applied problems on the basis of the constructed regulator are given. The first example demonstrates its use for managing an economic entity that is unstable in an open state. The second example contains the theoretical result in the form of a new algorithm for synthesizing a control system for a vector object of the "manipulation system" type, the physical model of which is the working organ of a one-bucket hydraulic excavator. The problem of constructing a control system for the process of stabilizing the bucket's angular position is considered. The algorithm for designing the control system and the corresponding control algorithm applied to the solution of the problem stabilizing the bucket of the excavator-robot along a given surface can be useful for any manipulation systems.

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ УДАЛЕННОГО ДИАГНОСТИРОВАНИЯ ТЕХНИЧЕСКОГО СОСТОЯНИЯ АГРЕГАТОВ

Д. Х. Имаев[1], **М. Л. Немудрук**[2], **М. С. Федоров**[3], **М. Ю. Шестопалов**[4]
(*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина), 197376, С.-Петербург, ул. Проф. Попова, 5, nemomax@mail.ru*)

Аннотация

Описаны результаты исследований по интеллектуализации процедур удаленного диагностирования агрегатов электродвигатель-винтовой насос, предназначенных для перекачки темных нефтепродуктов. Сравниваются два подхода к диагностированию кавитационных режимов винтовых насосов — явный алгоритм обработки данных датчика вибрации и обучение искусственных нейронных сетей.

Ключевые слова: винтовой насос, нефтепродукт, кавитация, диагностирование, нейронная сеть.

Введение

Непрерывно-периодические процессы перевалки и хранения нефтепродуктов включают операции приема нефтепродуктов в резервуарные парки, хранения и отгрузку. Особенностью технологических процессов перевалки нефтепродуктов является территориальная рассредоточенность оборудования — трубопроводной сети, насосов, резервуаров, задвижек, клапанов и других агрегатов, что ставит задачу удаленного диагностирования их технического состояния. Прием и отгрузка темных нефтепродуктов осуществляются винтовыми насосами (ВН), которые склонны к кавитационным режимам, вызывающим недопустимые вибрации и, как следствие, приводящим к преждевременному выходу из строя дорогостоящего оборудования [1—3]. В настоящее время отсутствуют специализированные датчики, доставляющие информацию о таких режимах ВН. Кавитационные режимы дистанционно детектируются оператором на основе интегрированного анализа данных от целого ряда датчиков. Субъективный характер анализа данных различными по квалификации операторами на практике приводит к задержкам в обнаружении неисправностей и приводит к ошибкам в принятии мер по их устранению. Необходимо исследовать и разработать методы и алгоритмы автоматизации обработки данных в задаче удаленного диагностирования технического состояния и режимов работы насосных агрегатов.

Предлагаемый доклад посвящен сравнению двух подходов к диагностированию кавитационных режимов винтовых насосов для перекачки темных нефтепродуктов — явный алгоритм обработки данных датчика вибрации и обучение искусственных нейронных сетей.

Управляемый процесс перевалки темных нефтепродуктов

Типичная железнодорожная эстакада слива темных нефтепродуктов (мазута марок М-40, М-100 и топлива Э-4) нефтяного терминала позволяет одновременно сливать десятки цистерн. На рис. 1 приведена технологическая схема процесса слива, где: Ц — цистерна; Н — насос; ТО — теплообменник; БЕ — буферная емкость; Ф — фильтр.

Процесс слива складывается из двух основных операций — разогрева нефтепродуктов в цистернах и перекачки в продуктовый резервуарный парк.

---

[1] Доктор технических наук, профессор кафедры АПУ СПб ГЭТУ «ЛЭТИ».
[2] Инженер, ассистент кафедры АПУ СПб ГЭТУ «ЛЭТИ» .
[3] Кандидат технических наук, доцент кафедры АПУ СПб ГЭТУ «ЛЭТИ»
[4] Доктор технических наук, доцент, заведующий кафедрой АПУ СПб ГЭТУ «ЛЭТИ».

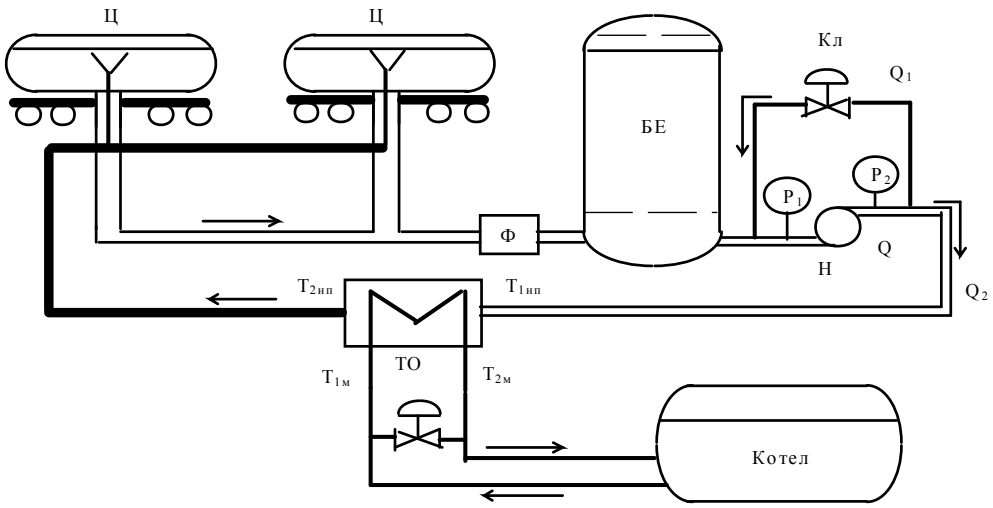


Рис. 1. Технологическая схема процесса слива

В холодный период года темные нефтепродукты должны быть разогреты и слиты за установленное время (от 4 до 10 часов в зависимости от погодных условий), для чего технологические сети оборудуются системами обогрева. Разогрев нефтепродуктов осуществляется циркуляционным способом (см. рис. 1). Мазут с установки нижнего слива, пройдя фильтры очистки от механических примесей и буферную емкость, поступает на один из насосов и далее в теплообменник. Разогретый до необходимой температуры мазут через телескопический гидромонитор вводится внутрь цистерны, где перемешивается с холодным.

Процесс перевалки характеризуется следующими основными переменными:

P_{out} — давление на выкиде насоса;

P_{ina} — давление на всасе насоса после фильтра;

P_{ini} — давление перед фильтром;

$T_{буф}$ — температура продукта в буферной емкости;

$T_{цист}$ — температуры на выходе из ж/д цистерн;

$T\text{то}_{out}$ —температура на выходе из теплообменников;

$T\text{эст}_{out}$ —температура на выходе из эстакады слива.

Предусмотрены контроль и регулирование температуры продуктов в трубопроводах и на входе/выходе теплообменников. В буферной емкости контролируются минимально и максимально допустимые уровни, при достижении которых насосы должны быть остановлены.

Возмущающими воздействиями на процесс являются исходная температура продукта из-за погодных условий на маршруте следования груза и температура окружающей среды. Кроме того, состояние основных технологических агрегатов на разных стадиях износа обеспечивают разную производительность и время срабатывания, что делает актуальным постоянный контроль технического состояния для своевременного обнаружения деградации основных параметров агрегатов. Прежде всего, это касается вибрации насосного агрегата, а также температуры подшипников двигателя насосного агрегата и валов насосов, температуры статора насоса и ток электродвигателя насоса.

При выборе схемы слива учитывается множество факторов. Для оптимизации управления операциями по перевалке нефтепродуктов — слива, перекачки, хранения и отгрузки — необходимо принять показатели, которые дают возможность объективно оценивать и оптимизировать технологическую схему (маршруты). Основной минимизируемый показатель – время слива (перевалки). Дополнительными показателями являются энергозатраты на разогрев и перекачку, расход ресурс технологических агрегатов, время работы в режимах близких к критическим и в критическом режиме, критическая скорость нарастания давления в теплообменниках.

Технологическая схема слива формируется оператором-технологом посредством системы дистанционного управления запорной электроприводной арматурой (электрозадвижки). После

того, как рассчитана и сформирована оптимальная технологическая схема слива, оператор-технолог запускает процесс циркуляции. В процессе реализации технологической схемы необходимо контролировать протекание процессов перевалки в условиях возмущений с помощью управляющих воздействий: частоты вращения вала насоса N; положения клапана подачи теплоносителя в теплообменник Pos_{Clap}. При управлении процессом слива должны минимизироваться время операции, энергозатраты и расход ресурса оборудования с учетом ограничений диапазона допустимых скоростей вращения вала, производительности теплообменников, давления в трубопроводах и др. Одновременно необходимо контролировать состояние технологической схемы (положение запорной арматуры) и агрегатов (недопустимые режимы работы оборудования).

Автоматизация управления процессом перевалки нефтепродуктов

Сложная система управления процессом перевалки нефтепродуктов образована совокупностью подсистем, взаимодействие которых способно обеспечивать выполнение собственных функций и достижение общесистемных целей [4]. Каждая подсистема управляет частью технологического объекта или объединяет технические средства, выполняющие определенную функцию. Для автоматизации управления таким сложным объектом строятся децентрализованные системы, состоящие из множества контроллеров, интегрированных в единую систему посредством локальных вычислительных сетей. Система управления является следствием декомпозиции алгоритмов контроля и управления, как по горизонтали, так и по вертикали. Горизонтальное разбиение функций алгоритмов контроля и управления должно обеспечивать, по возможности, максимальную функциональную и территориальную децентрализацию и минимальный интерфейс между подфункциями и реализующими их блоками [5], [6].

АСУ ТП нефтяного терминала создается как территориально распределенная и функционально децентрализованная система, связанная с информационно-программной системой оперативной службы. По своей организации АСУ ТП является иерархической двухуровневой децентрализованной системой. Верхний уровень системы представляет собой операторские станции управления основным и вспомогательными технологическими процессами. Нижний уровень системы представлен локальными управляющими станциями, построенными на базе программируемых компьютерных контроллеров, осуществляющих функции сбора, первичной обработки данных и локального управления оборудованием. Концепция реализации АСУ ТП в виде двухуровневой системы определяет необходимость сопряжения локальных контуров управления всеми процессами, которые размещены в аппаратуре «низовой» автоматики, и управляющих централизованных функций оператора.

Подсистемы АСУ ТП целесообразно выделять в соответствии с разбиением на группы технологического оборудования (см. рис. 1).

Подсистема управления железнодорожными эстакадами слива нефтепродуктов обеспечивает:

- визуализацию технологической схемы эстакад слива на операторской станции диспетчера;
- контроль и индикацию состояния задвижек эстакад;
- опрос и индикацию параметров с КИП;
- дистанционное управление задвижками по командам диспетчера;
- блокировки при ошибочных действиях диспетчера;
- аварийную сигнализацию;
- регистрацию действий диспетчера при дистанционном управлении и действий полевого персонала при местном управлении.

Каждое сливное устройство снабжено датчиками температуры слива из каждой цистерны и выходе из приемного коллектора. На эстакаде целесообразно использовать электрозадвижки с датчиками положения, позволяющими регулировать интенсивность подачи разогретого мазута в цистерну и скорость отбора разогретого топлива.

Подсистема дистанционного управления запорной арматурой технологических трубопроводов обеспечивает: визуализацию технологической схемы трубопроводов и резервуарных парков на операторской станции; контроль и индикацию состояния задвижек; дистанционное

управление задвижками по командам диспетчера терминала; блокировки при ошибочных действиях диспетчера; регистрацию действий диспетчера при дистанционном управлении и действий полевого персонала при местном управлении. Практически все технологические схемы работы с трубопроводами, резервуарными парками и насосными станциями, формируются диспетчерами дистанционно.

Диспетчеры должны быть полностью информированы о состоянии оборудования, так как недостаток своевременной информации при дистанционном управлении может приводить к аварийным ситуациям и материальным потерям.

Важным является фактор доверия диспетчеров к системе дистанционного управления задвижками. Необходимо повышать степень автоматизации запорной арматуры. Электрприводы задвижек должны удовлетворять ряду требований по информационным и управляющим сигналам:

1 — задвижка закрыта (сигнал с конечного выключателя);

2 — задвижка открыта (сигнал с конечного выключателя);

3 — стоп (сигнал с кнопки “Стоп” местного пульта управления);

4 — движение на открытие (сигнал с магнитного пускателя);

5 — движение на закрытие (сигнал с магнитного пускателя);

6 — дистанционный режим (сигнал с переключателя «Местный/ Дистанционный» местного пульта управления);

7 — местный режим (сигнал с переключателя «Местный/Дистанционный» местного пульта управления);

8 — авария (сигнал с выключателей муфт предельного момента);

9 — степень открытия задвижки.

Дистанционных управляющих сигналов достаточно двух — для схемы управления потенциальными сигналами или трех — для схемы управления импульсными сигналами (1 — движение на открытие; 2 — движение на закрытие; 3 — стоп). Большое количество информационных и управляющих сигналов с задвижек приводит к использованию множества модулей ввода/вывода дискретных и аналоговых сигналов.

Современным вариантом дистанционного управления запорной арматурой является применение задвижек с «интеллектуальным» приводом, снабженным собственной системой локального контроля и управления. Встроенный микроконтроллер осуществляет сбор данных с конечных выключателей, датчика положения задвижки, контактов состояния пускателя и других датчиков, управляет включениями и выключениями задвижки. Пускатель, тепловая защита и другое электрооборудование расположены в корпусе привода. Для взаимодействия с такими приводами используются специализированные промышленные сети. Связь с центральным управляющим контроллером осуществляется по двухпроводной цифровой магистрали. К одной магистрали подключаются десятки «интеллектуальных» задвижек, что снимает необходимость в шкафах с силовым электрооборудованием приводов, уменьшает число входных/выходных модулей локальных станций, сокращает длины силовых кабелей, упрощает и удешевляет электромонтаж.

Подсистема контроля и управления котельными должна обеспечивать:

- обмен данными между операторскими станциями диспетчера терминала и котельных;
- визуализацию технологических параметров и состояния оборудования котельных на операторских станциях диспетчера терминала и оператора котельных;
- дистанционное управление режимами работы котлов;
- аварийную сигнализацию.

Предполагается, что котельная оснащена системой управления на базе программируемых логических контроллеров и операторской станцией. Кроме того, система управления должна быть открытой, т. е. должны быть предусмотрены аппаратные и программные средства для информационного обмена с АСУ ТП. Эти средства должны позволять оператору котельных и диспетчеру контролировать работу котельной, а также изменять режим работы котельной путем изменения задания управляющих уставок.

Подсистема контроля и управления технологической насосной обеспечивает:

- визуализацию схемы насосной на операторской станции диспетчера;

- контроль и индикацию состояния насосов и задвижек, данных контрольно-измерительных приборов (КИП);
- дистанционное управление насосами по командам диспетчера;
- опрос и индикацию параметров с КИП;
- блокировки ошибочных действий диспетчера;
- технологические блокировки;
- аварийную сигнализацию;
- регистрацию действий диспетчера при дистанционном управлении и действий полевого персонала при местном управлении.

Основной задачей подсистемы является дистанционное управление технологическими насосами и контроль их состояния. Для обеспечения бесперебойной работы и сигнализации аварийных ситуаций насосы оборудуются датчиками температуры масла в подшипниках вала насоса и датчиками измерения давления. Операция разогрева мазута в цистернах на сливной эстакаде предполагает плавную регулировку производительности насосов в зависимости от температуры мазута. Для этого мазутные циркуляционные насосы предлагается оснащать приводом с частотными преобразователями, позволяющими регулировать частоту вращения в определенных пределах.

Удаленное диагностирование работы винтовых насосов

Прием и отгрузку темных нефтепродуктов осуществляют винтовые насосы (ВН), которые склонны к кавитационным режимам, вызывающим недопустимые вибрации и, как следствие, приводящим к преждевременному выходу из строя дорогостоящего оборудования [1]—[3]. В настоящее время отсутствуют специализированные датчики, доставляющие информацию о таких режимах ВН.

Для обеспечения бесперебойной работы и сигнализации аварийных ситуаций насосы оборудуются датчиками температуры масла в подшипниках вала насоса и датчиками измерения давления. Переменные, характеризующие состояние ВН: $T_{b1, \ldots, b6}$ — температуры подшипников двигателя насосного агрегата и валов насосов; $Ts_{1, 2, 3}$ — температура статора насоса; ток нагрузки электродвигателя насоса; V_{bpump} — вибрация насосного агрегата. Возникновение и интенсивность вибрации зависят от частоты вращения двигателя, вязкости перекачиваемого продукта, которая в свою очередь зависит от температуры и технического состояния узлов насосного агрегата.

В настоящее время кавитационные режимы дистанционно детектируются оператором на основе интегрированного анализа данных от целого ряда датчиков — вибрации, давления на входе и выходе насоса, частоты вращения и тока нагрузки двигателя, температуры перекачиваемого продукта и др. Субъективный характер анализа данных различными по квалификации операторами на практике приводит к задержкам в обнаружении неисправностей и приводит к ошибкам в принятии мер по их устранению. Необходимо исследовать и разработать методы и алгоритмы автоматизации обработки данных в задаче удаленного диагностирования технического состояния и режимов работы насосных агрегатов.

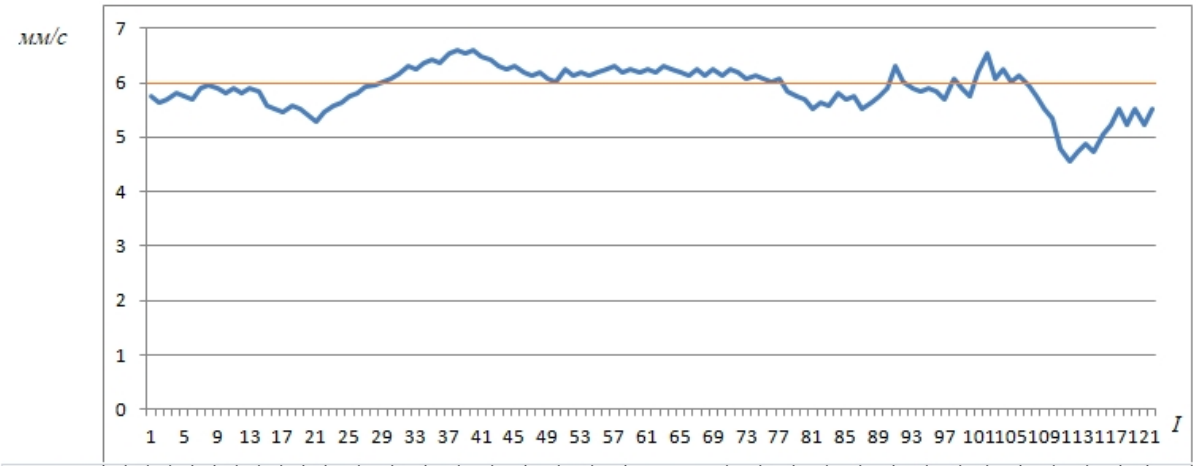
Можно указать на два подхода к решению задачи: 1 – программирование четко описанных («явных») алгоритмов обработки первичных данных от штатных датчиков; 2 – применение методов искусственного интеллекта (машинного обучения).

Явный алгоритм обработки данных для диагностирования кавитационных режимов использует данные установленного на корпусе насосного агрегата акселерометра, измеряющего мгновенные значения виброскорости. Алгоритм «виртуального» датчика кавитаций регистрирует превышения порогового уровня вибрации определенной длительности за диагностический интервал времени. По количеству превышений делается вывод о наступлении события «Кавитация насосного агрегата».

При современном подходе с использованием датчика виброскорости происходит регистрация факта превышения аварийного значения виброскорости. Вместе с тем, единичное превышение порога само по себе не означает, что на ВН имеют место кавитационные процессы. При приближении к кавитационному режиму количество «выбросов» в показаниях вибродатчиков

растет. Растет также количество записей в журнале регистрации аварийных ситуаций. Каждое такое событие требует действий со стороны оператора технолога — аварийное сообщение должно им квитироваться и должны вноситься коррективы в технологический процесс (изменяться производительность насоса, корректироваться подача теплоносителя в теплообменники и др.). Но из опыта эксплуатации оператор понимает, что такое единичное событие как однократный выход значения виброскорости за допустимые пределы никаких корректирующих мероприятий не требует. Фактически от оператора-технолога требуется провести анализ комплекса показателей (частота возникновения вибрационных событий либо полный уход их показаний в «красную зону», значение давления на всасе насоса, температура перекачиваемого продукта) для обоснованного заключения о наличии кавитации по совокупности показаний нескольких датчиков и их изменении во времени. Помимо сложности такого анализа имеется проблема, связанная с тем, что оператор теряет в определенной мере доверие к показаниям системы обнаружения вибраций, поскольку значительная часть аварийных сообщений фактически таковыми не являются. К тому же количество таких сообщений от разных насосов может быть значительным.

Чтобы избавить оператора от несвойственных ему функций, внедрен алгоритм, основанный на следующей идее обнаружения опасных вибрационных режимов. Вводятся понятия «вибрационное событие» и «вибрационное событие вследствие кавитационных режимов». С определенной периодичностью (около десяти раз в секунду) измеряется виброскорость. Вводится окно времени за которое фактически проводится анализ выборки. Формируется вектор булевых переменных, каждый элемент которого кодирует результат сравнения показаний датчика с пороговым значением: 0, если показание меньше порога; 1, если больше. На каждом цикле добавляется новое показание, а самое старое — отбрасывается. На рис. 2 показан пример формирования вектора (массива) состояний.



Индекс	0	1	2	3	4	5	6	7	8	9		24	25	26	27	28	29	30	31		75	76	77	78	79	80	81		90	91	92	93		120
Значение	0	0	0	0	0	0	0	0	0	0	...	0	1	1	1	1	1	1	1	...	1	1	1	1	0	0	0	...	0	1	1	0	...	0

Рис. 2. Соответствие графика виброскорости массиву состояний/

Подсчитывается, когда виброскорость превышала пороговое значение в течении *n* отсчетов (*n* — «глубина» фильтрация для исключения единичных выбросов, связанных, например с наводками). По числу превышений порогового значения делается вывод о начале вибрационного события. Дополнительный анализ показаний датчика давления на всасе насоса позволяет заключить о наступлении: «Вибрационное событие кавитационной природы», либо — «Вибрационное событие».

Алгоритм выявления кавитационного режима ВН состоит в выполнении следующей последовательности операций (рис. 3):

1 — формирование массива событий с учетом выбранной глубины;

2 — сравнение текущего значения виброскорости с пороговым;

3 — сдвиг значений в массиве событий на единицу влево;

4 — запись результата сравнения (1 — превышение; 0 — норма) в ячейку массива событий;
5 — подсчет количества превышений порогового значения при условии длительности сигнала превышения выше *n* — числа событий превышения порогового уровня вибрации;
6 — время превышения аварийного порога «Кавитация»; иначе — «Вибрация вследствие технического состояния ВН».

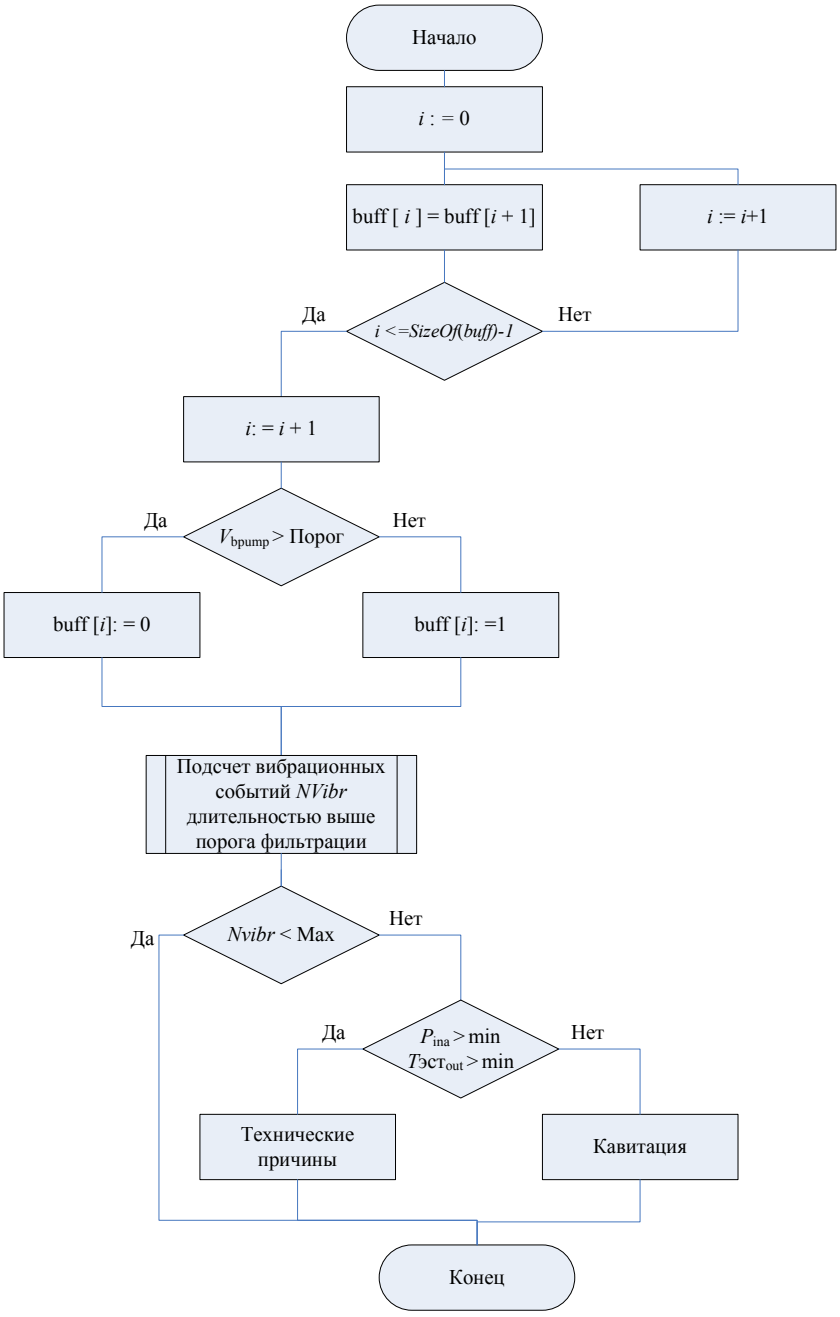


Рис. 3. Блок-схема явного алгоритма

На рис. 3 приняты следующие обозначения: *buff* – массив состояний, *Nvibr* – количество превышений пороговых значений за период, *i* – индекс массива состояний.

Интеллектуальный метод диагностирования.
Диагностирование ВН следует рассматривать как задачу отнесения его технического состояния к определенному классу в выбранном пространстве диагностических признаков (ДП). Современные интеллектуальные технологии предоставляют мощные средства

классификации данных, комбинирующие Neuro- и Fuzzy-технологии с целью рационального сочетания: 1 — способности искусственных нейронных сетей к обучению на верифицированных данных; 2 — возможности формализации знаний экспертов в виде баз правил механизмов вывода в нечеткой логике [7]—[10].

Обнаружение кавитационных режимов ВН связано с последовательной обработкой данных (рис. 4). Обработка данных измерений $y_q[k]$ формирует ДП d_i [K], которые подаются на входы нейронных классификаторов NNi, обученных выявлять симптомы c_j [K] потенциальных неисправностей. Выявление симптома означает обнаружение неисправности, причем, только в частном случае единственный симптом будет означать и диагноз D [K]. Как правило, диагноз ставится по совокупности симптомов $\{ c_j; j = 1, 2, \ldots\}$.

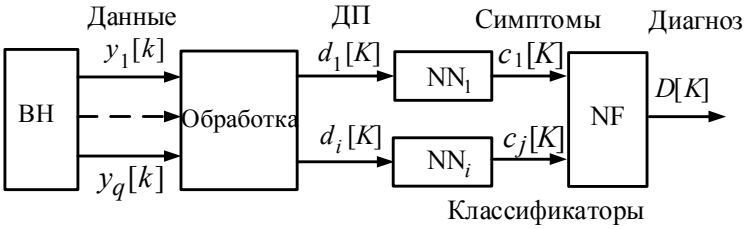


Рис. 4. Последовательность обработки данных

Предлагается ввести понятие ДП различных категорий [10]. *ДП нулевой категории* $d^0[k]$ — мгновенные значения переменных $y_q[t]$, т. е. числовые последовательности $y_q[k] \equiv y(t = kT)$; $k = 1, 2, \ldots$, где T — период дискретизации времени. Нарушение переменными ВН допустимых границ, безусловно, свидетельствует о неисправностях. ДП нулевой категории $d^0[k]$ обеспечивают примитивный уровень диагностирования, достоинство которого заключается в максимальной скорости обнаружения неисправностей. Недостаток ДП нулевой категории $d^0[k]$ — высокая чувствительность к помехам в случае четких пороговых устройств. За *ДП первой категории* $d^1[K]$ — отрезки дискретных последовательностей (случайные выборки) $\{ y_q[k]; k = 1, 2, \ldots, K\}$, где K — длина выборки. *ДП второй категории* $d^2[K]$ — результаты статистической обработки: средние, дисперсии с выборки, данные спектрального или корреляционного анализа. Особенность ДП заключается в том, что сигналы доставляют информацию с периодичностью KT, определяемой длиной выборки K и периода дискретизации T непрерывных процессов, что приводит к задержке обнаружения неисправностей.

Принимается концепция «одна неисправность — один классификатор симптома», т. е. для неисправности «кавитационный режим ВН» обучается своя нейронная сеть. Процедура обучения нейроклассификатора, выявляющего симптом и ставящего диагноз о кавитационном режиме работы ВН, складывается из следующих основных операций.

1. Подготовка обучающих последовательностей. Операция представляет особую трудоемкость и выполняется совместно с операторами. Использование и анализ архивных данных штатных АСУ ТП предполагает автоматизацию процедур формирования обучающих последовательностей с учетом необходимости «синхронизации» данных измерений с оборудования (входы ИНС) и данных о действиях разных операторов (выходы ИНС). Следует выбрать максимально информативные для выявления симптома неисправности c_j [K] данные измерений — переменные ВН $y_q[k]$. Выявлено, что такими данными являются показания акселерометра, измеряющего мгновенные значения виброскорости. Возникновение и интенсивность вибрации зависят от частоты вращения двигателя, вязкости перекачиваемого продукта, которая в свою очередь зависит от температуры и технического состояния узлов насосного агрегата. Предлагается ДП первой категории $d^1[K]$ — отрезки дискретных последовательностей, когда вход ИНС — скалярная функция векторного аргумента размерности $K \times 1$.

2. Кодирование выходных обучающих последовательностей сети (например, «Нормальный режим» — 1, «Кавитация» — 0).

3. Выбор архитектуры и тренировка нейронной сети проводятся с помощью специальной программы nprtool инструментального комплекса *MATLAB/Neural Network Toolbox*. Выбрана двухслойная сеть прямого действия с 20 нейронами в скрытом слое.

Пример диагностирования технического состояния винтовых насосов для перекачки мазутов с помощью нейронной сети. На рис. 5 изображен результат обучения сети на языке графического редактора *Simulink*. Блок Display1 показывает нечеткую границу между нормальным и кавитационным режимами ВН. Блок Display1, по существу, дает *степень принадлежности* работы ВН к исправным режимам.

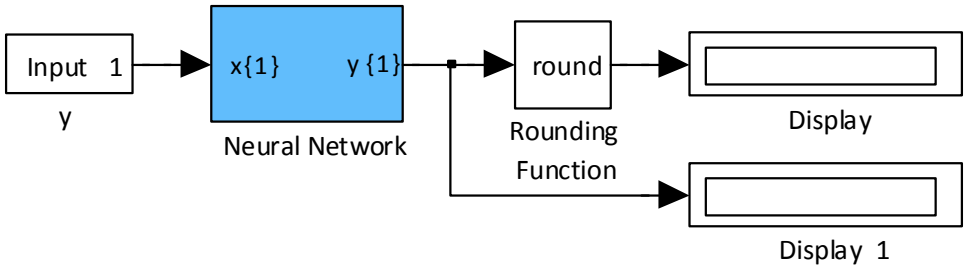


Рис. 5. Нейронная сеть, распознающая режим кавитации ВН

Таким образом, нейросетевой классификатор играет роль «фаззификатора» и может служить как вход последующего механизма вывода на нечеткой логике, ставящего диагнозы (см. рис. 4). Поскольку граница между нормальным и кавитационным режимами размыта, для получения четкой границы между режимами к выходу сети подключен блок округления round (см. рис. 5).

Заключение

Общей тенденцией отраслей нефтегазового сектора экономики является стремление к интеллектуализации систем контроля и управления, что объясняется территориальной рассредоточенностью трубопроводных сетей, насосов, компрессоров, резервуаров, задвижек, клапанов и др., что ставит актуальную задачу удаленного диагностирования агрегатов. Вместе с тем, для целого ряда интегральных показателей технического состояния агрегатов нет датчиков, данные которых позволяют обнаруживать и локализовать неисправности или недопустимые режимы функционирования. Примером является детектирование и идентификация интенсивности кавитационных режимов винтовых насосов для перекачки темных нефтепродуктов.

Задача решается путем создания «виртуальных» датчиков — различных способов обработки имеющихся данных с целью диагностирования агрегатов. Иллюстрируются два подхода к удаленному диагностированию агрегатов: 1 — программирование четко описанных («явных») алгоритмов обработки первичных данных от штатных датчиков; 2 — применение методов искусственного интеллекта. Явные алгоритмы решают задачу диагностирования с заложенной точностью в течение срока эксплуатации, однако при изменении постановки задачи необходимо менять и алгоритм. В докладе предложен один из явных алгоритмов детектирования кавитационных режимов винтовых насосов.

Особенностью применения нейронных классификаторов распознавания образов, т. е. интеллектуализации процедур диагностирования, является гибкость. В зависимости от обучающих последовательностей в виде различных диагностических признаков удается решать разные задачи. Приведен пример двухслойной нейронной сети прямого действия с двадцатью нейронами в скрытом слое, позволяющей диагностировать степень принадлежности режима работы насоса к кавитационному. Следует отметить трудоемкость формирования диагностических показателей на основе субъективных экспертных знаний. В докладе приводится простейший способ подготовки обучающих последовательностей: на входе сети в форме выборок «сырых» данных штатных датчиков виброскорости и на выходе — в виде действий нескольких операторов, управляющих сливом мазута из железнодорожных цистерн.

Предлагаемый подход к интеллектуализации удаленного диагностирования технического состояния агрегата нефтяного терминала, аппроксимирует действия «усредненного» оператора, что должно способствовать объективизации заключений о состоянии оборудования.

ЛИТЕРАТУРА

1. Картошкин М.Д. Топливоподача тепловых электростанций. — М.; Л.: Госэнергоиздат, 1961. 208 с.
2. Ляндо И.М. Эксплуатация мазутного хозяйства котельной промышленного предприятия. М.: Энергия, 1968. 151 с.
3. Назмеев Ю.Г. Мазутные хозяйства ТЭС. — М.: МЭИ, 2002. 612 с.
4. Имитационное моделирование производственных систем/ Под ред. А.А. Вавилова. — М.: Машиностроение. 416 с.; Berlin: Veb Verlag Technik, 1983. S. 380.
5. Автоматизированное управление технологическими процессами. Учебное пособие/ Зотов Н.С., Назаров О.В., Петелин Б.В. и др./ Под ред. В.Б. Яковлева. — Л.: Изд-во Ленингр. ун-та, 1988. 221 с.
6. Антонов В.Н., Наседкин А.В., Федоров М.С. Технические средства микропроцессорных распределенных систем управления. — Л.: ЛЭТИ, 1988. 72 с.
7. Ayoubi M., Isermann R. Neuro-fuzzy systems for diagnosis// Fuzzy Sets and Systems, No. 89, 1997. P. 289–307.
8. Krebs V., Korablev J., Shestopalov M. Neuro-Fuzzy-technology based diagnostics in technical control systems // Межд. конф. по мягким вычислениям и измерениям, Санкт-Петербург, 17–19 июня, 2004 г., т. 2. С. 208–210.
9. Шестопалов М.Ю., Имаев Д.Х. Диагностирование систем управления с помощью нейросетевых классификаторов распознавания неисправностей// Сб. докладов XVI Междунар. конф. по мягким вычислениям и измерениям SCM-2013, Изд-во СПбГЭТУ «ЛЭТИ», 23–25 мая, т. 1, 2013. С. 109–111.
10. Шестопалов М.Ю. Системы отказоустойчивого управления технологическими процессами. — СПб, Изд-во «Элмор», 2013. 308 с.

D.H. Imaev, M.L. Nemudruk, M.S. Fyodorov, M.Yu. Shestopalov (Saint Petersburg State Electrotechnical University «LETI»)

An Intellectualization of Systems for Pumping Units Remote Diagnostics

The results of research on the intellectualization of procedures for remote diagnosis of screw pump units for dark petroleum products are described. Compares two approaches to the diagnosis of cavitation modes screw pumps — regular algorithm for processing data of the vibration sensor and the training of artificial neural networks.

Секция 5
ОБРАБОТКА СИГНАЛОВ В СИСТЕМАХ УПРАВЛЕНИЯ И СВЯЗИ (СЕКЦИЯ ПАМЯТИ А. Е. БАРАБАНОВА)

О ПЕРИОДИЧЕСКИХ В СРЕДНЕМ РЕШЕНИЯХ ЛИНЕЙНЫХ НЕОДНОРОДНЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ СО СЛУЧАЙНЫМИ КОЭФФИЦИЕНТАМИ

В. Г. Задорожний[1]
(*Воронежский государственный университет,
Воронеж, 394018, Университетская пл., 1, zadorozhny@amm.vsu.ru*)

Г. А. Курина[2]
(*Воронежский государственный университет, Воронеж, 394018, Университетская пл., 1;
Воронежский экономико-правовой институт, Воронеж, 394000, Ленинский пр-т, 119-А;
Федеральный исследовательский центр "Информатика и управление" Российской академии наук,
Москва, 119333, ул. Вавилова, 44/2, kurina@math.vsu.ru*)

Аннотация

В докладе будут представлены новые результаты, касающиеся определения математического ожидания для решений линейного неоднородного уравнения диффузии с одной пространственной переменной и случайными коэффициентами. Для этого получено вспомогательное не случайное дифференциальное уравнение с обычными и вариационными производными и детерминированное начальное условие.

Для частного случая рассматриваемого уравнения, а именно уравнения теплопроводности, в котором коэффициент перед неизвестной функцией является гауссовым или равномерно распределенным, а также для обыкновенного неоднородного скалярного линейного дифференциального уравнения первого порядка со случайными коэффициентами приводятся условия существования периодических в среднем решений.

Если рассматривать неоднородность в качестве управляющей функции, то полученные результаты можно трактовать как нахождение управления, обеспечивающего периодичность математического ожидания решения.

Ключевые слова: линейные дифференциальные уравнения со случайными коэффициентами, вариационная производная, математическое ожидание решения и его периодичность

Введение

Хорошо известны (см., например, [1], с. 90-98) условия существования периодического решения обыкновенного линейного дифференциального уравнения при условии, что коэффициент при искомой функции и правая часть являются детерминированными периодическими функциями. Задача существенно усложняется, если коэффициенты являются случайными процессами.

Приведем здесь некоторые результаты из [2,3], касающиеся условий, обеспечивающих периодичность математического ожидания и дисперсионной функции решений скалярных обыкновенных линейных неоднородных дифференциальных уравнений первого порядка со случайными коэффициентами.

Решение дифференциального уравнения со случайными коэффициентами называется периодическим в среднем, если его математическое ожидание является периодической функцией. Предположим, что коэффициент при неизвестной и правая часть являются независимыми случайными процессами, заданными характеристическими функционалами. При этом не предполагается представление коэффициентов в виде суммы детерминированной функции и белого шума.

Далее используется понятие вариационной производной. Напомним её определение (см., например, [4], с. 14).

[1] Доктор физико-математических наук, профессор.
[2] Доктор физико-математических наук, профессор.

Пусть y – функционал на пространстве $L_1(\mathbb{R}_+)$ комплекснозначных функций, интегрируемых на $\mathbb{R}_+$, т. е. $y : L_1(\mathbb{R}_+) \to \mathbb{C}$, и $v \in L_1(\mathbb{R}_+)$, $h \in L_1(\mathbb{R}_+)$.

Если

$$y(v+h) - y(v) = \int_{\mathbb{R}_+} \varphi(t,v)h(t)dt + o(h),$$

где интеграл понимается в смысле Лебега и является линейным ограниченным относительно h функционалом на $L_1(\mathbb{R}_+)$, а $o(h)$ обозначает бесконечно малую высшего порядка относительно h, то отображение $\varphi : \mathbb{R}_+ \times L_1(\mathbb{R}_+) \to \mathbb{C}$ называется вариационной производной функционала y в точке v и обозначается $\frac{\delta y(v)}{\delta v(t)}$.

Для нахождения математического ожидания решения сначала строится вспомогательное детерминированное дифференциальное уравнение первого порядка с обычной и вариационной производными и с детерминированным начальным условием. Выясняется, что это уравнение интегрируется, т. е. получается формула для общего решения. Определяются условия, при которых вспомогательное уравнение имеет периодическое решение, находится формула для такого решения.

Чтобы определить вторую моментную функцию решения, вводится новое вспомогательное отображение. Для этого отображения получается детерминированное дифференциальное уравнение, содержащее обычную производную и вариационные производные. Также находится детерминированное начальное условие для решения уравнения. Однако полученная задача имеет бесчисленное множество решений и возникает проблема выбора соответствующего решения. Поскольку вторая моментная функция решения является симметричной функцией двух переменных, то ищется симметричное по этим переменным решение вспомогательной задачи. Такое решение оказывается единственным, и оно выписывается в явном виде. Анализ вида решения позволяет найти условия, при которых решение вспомогательного уравнения оказывается периодическим. Из этого решения находится формула для периодической второй моментной функции. Так как вторая моментная функция симметрична, то получается периодическая вторая моментная функция по обеим переменным. С помощью формулы для второй моментной функции получается вид дисперсионной функции.

Приводятся условия, при которых формула для периодического математического ожидания решения совпадает с формулой периодического решения соответствующего детерминированного уравнения.

Решение дифференциального уравнения со случайными коэффициентами называется периодическим в широком смысле, если его математическое ожидание и дисперсионная функция являются периодическими с одинаковыми периодами.

При условии, что коэффициент при искомой функции является гауссовым случайным процессом или имеет равномерное распределение, в [2,3] получены условия существования периодического в среднем решения и явная формула для периодического математического ожидания. Для этих случаев найдены условия существования периодических в широком смысле решений и формула для периодической дисперсионной функции. Рассмотрен также и более сложный резонансный случай, когда линейное однородное уравнение со случайным коэффициентом допускает ненулевое периодическое в среднем решение.

В докладе рассматривается задача Коши для неоднородного линейного дифференциального уравнения диффузии с одной пространственной переменной, коэффициенты которого являются статистически независимыми случайными гауссовыми процессами. Предполагается, что случайные процессы заданы характеристическими функционалами.

Получено вспомогательное не случайное дифференциальное уравнение с обычными и вариационными производными и детерминированное начальное условие. Найдены явное представление для обобщенного решения вспомогательной задачи и явный вид математического ожидания обобщенного решения исходной задачи для уравнения диффузии.

Для частного случая рассматриваемого уравнения, а именно уравнения теплопроводности, приводятся условия существования периодических в среднем решений.

Математическое ожидание решений линейного уравнения второго порядка с частными производными и случайными коэффициентами

Постановка задачи. Рассмотрим задачу Коши

$$\frac{\partial u}{\partial t} = \varepsilon_1(t,\omega)\frac{\partial^2 u}{\partial x^2} + \varepsilon_2(t,\omega)u + f\ (t,x,\omega), \tag{1}$$

$$u(0,x) = u_0(x,\omega), \tag{2}$$

где $t \in \mathbb{R}_+$, $x \in \mathbb{R}$ и $\varepsilon_1 \geq 0$, ε_2, f, u_0 - случайные процессы, ω - случайное событие. Если ε_1, ε_2, f, u_0 не являются случайными процессами, то решение задачи известно (см., например, [5], с. 64 и [6], с. 218). Если $\varepsilon_1 \geq 0$, ε_2, f, u_0 – случайные процессы, то решение также является случайным процессом. Мы не ставим себе задачей определить функцию распределения либо плотность распределения, либо характеристический функционал решения - эти задачи пока достаточно сложны. Мы находим математическое ожидание $M(u(t,x))$ решения задачи (1), (2). Для этого применяем метод, основанный на сведении задачи к не случайному дифференциальному уравнению с обычными и вариационными производными (см. [4]).

Сведение к детерминированной задаче. Пусть $L_1(\mathbb{R}_+ \times \mathbb{R})$ – пространство суммируемых на $\mathbb{R}_+ \times \mathbb{R}$ функций. Введем обозначение

$$e(v_1,v_2,z) = exp\left(i\int_{\mathbb{R}_+}[\varepsilon_1(s)v_1(s) + \varepsilon_2(s)v_2(s)]ds + i\int_{\mathbb{R}}\int_{\mathbb{R}_+} f(s,\tau)z(s,\tau)\,ds\,d\tau\right).$$

где $i = \sqrt{-1}$, $v_1, v_2 \in L_1(\mathbb{R}_+)$, $z \in L_1(\mathbb{R}_+ \times \mathbb{R})$. Будем предполагать, что известен характеристический функционал (см., например, [4], с. 30) $\psi(v_1,v_2,z) = M(e(v_1,v_2,z))$ случайных процессов ε_1, ε_2, f. Здесь математическое ожидание M вычисляется по функции распределения процессов $\varepsilon_1, \varepsilon_2$, f.

Умножим уравнение (1) и условие (2) на $e(v_1,v_2,z)$ и вычислим математическое ожидание полученных выражений

$$M\left(\frac{\partial u}{\partial t}e(v_1,v_2,z)\right) = M\left(\varepsilon_1(t,\omega)\frac{\partial^2 u}{\partial x^2}e(v_1,v_2,z)\right)$$
$$+M\big(\varepsilon_1(t)ue(v_1,v_2,z)\big) + M\big(f(t,x,\omega)e(v_1,v_2,z)\big), \tag{3}$$

$$M\left(u(0,x)e(v_1,v_2,z)\right) = M\left(u_0(x)e(v_1,v_2,z)\right). \tag{4}$$

В дальнейшем зависимость случайных процессов от случайного события ω в записи не отражается и предполагается, что u_0 не зависит от случайных процессов ε_1, ε_2, f.

Введем отображение

$$y(t,x,v_1,v_2,z) = M\big(u(t,x)e(v_1,v_2,z)\big).$$

Отметим, что $y(t,x,0,0,0) = M\big(u(t,x)\big)$. Равенства (3), (4) (формально) можно записать с помощью y

$$\frac{\partial y}{\partial t} = -i\frac{\partial^2}{\partial x^2}\frac{\delta_p y}{\delta v_1(t)} - i\frac{\delta_p y}{\delta v_2(t)} - i\frac{\delta_p \psi}{\delta z(t,x)}, \tag{5}$$

$$y(0,x,v_1,v_2,z) = M(u_0(x))\psi(v_1,v_2,z). \tag{6}$$

Здесь, например, $\frac{\delta_p y}{\delta v_1(t)}$ – частная вариационная производная по переменной v_1. Задача (5), (6) является детерминированной (не зависит от случайного события ω), и содержит обычные и вариационные производные. Задача получена формально, но естественным является следующее.

Определение. Математическим ожиданием $M(u(t,x))$ решения задачи (1), (2) называется $y(t,x,0,0,0)$, где $y(t,x,v_1,v_2,z)$ – решение задачи (5), (6).

Решение задачи (5), (6). Пусть процессы ε_1, ε_2, f независимы, являются гауссовыми и заданы характеристическими функционалами

$$\varphi_{\varepsilon_1}(v_1) = exp\left(i\int_{\mathbb{R}_+} a_{\varepsilon_1}(s)v_1(s)ds - \frac{1}{2}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} b_{\varepsilon_1}(s_1,s_2)v_1(s_1)v_1(s_2)ds_1ds_2\right),$$

$$\varphi_{\varepsilon_2}(v_2) = exp\left(i\int_{\mathbb{R}_+} a_{\varepsilon_2}(s)v_2(s)ds - \frac{1}{2}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} b_{\varepsilon_2}(s_1,s_2)v_2(s_1)v_2(s_2)ds_1ds_2\right),$$

$$\varphi_f(z) = exp(i\int_{\mathbb{R}_+}\int_{\mathbb{R}} a_f(s,\xi)z(s,\xi)dsd\xi$$

$$-\frac{1}{2}\int_{\mathbb{R}_+}\int_{\mathbb{R}}\int_{\mathbb{R}_+}\int_{\mathbb{R}} b_f(s_1,\xi_1,s_2,\xi_2)z(s_1,\xi_1)\,z(s_2,\xi_2)ds_1ds_2),$$

где

$$a_{\varepsilon_1}(s) = M(\varepsilon_1(s)) \geq 0, a_{\varepsilon_2}(s) = M(\varepsilon_2(s)),\ a_f(s,\xi) = M(f(s,\xi)),$$

$$b_{\varepsilon_1}(s_1,s_2) = M(\varepsilon_1(s_1)\varepsilon_1(s_2)) - M(\varepsilon_1(s_1))M(\varepsilon_2(s_2)),$$

$$b_{\varepsilon_2}(s_1,s_2) = M(\varepsilon_2(s_1)\varepsilon_2(s_2)) - M(\varepsilon_2(s_1))M(\varepsilon_2(s_2)),$$

$$b_f(s_1,\xi_1,s_2,\xi_2) = M(f(s_1,\xi_1)f(s_2,\xi_2)) - M(f(s_1,\xi_1))M(f(s_2,\xi_2)).$$

Решение задачи (5), (6) будем искать в пространстве обобщенных функций. Пусть $\theta(t) = 1$ при $t \geq 0$ и $\theta(t) = 0$ при $t < 0$,

$$\Phi_1(t,x) = \frac{\theta(t)}{2\sqrt{\pi t}}\exp(-\frac{x^2}{4t})$$

– фундаментальное решение уравнения теплопроводности $\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$ (см., например, [6], с. 191) и $\Phi_2(t,x)$ – фундаментальное решение уравнения $\frac{\partial u}{\partial t} = \frac{\partial^4 u}{\partial x^4}$. Пусть δ – дельта-функция, $*$ – знак свертки по переменным x, t. Определим функцию $\chi(\tau) = \chi(s,t,\tau)$ переменной τ следующим образом: $\chi(\tau) = sign(\tau - s)$, если τ принадлежит отрезку с концами s, t, и равна нулю в противном случае.

В дальнейшем нам потребуется следующая

Теорема 1. ([4], с. 166) Пусть a – непрерывная функция на $\mathbb{R}_+$ и функционал y: $L_1(\mathbb{R}_+) \to \mathbb{C}$ имеет вариационную производную $\frac{\delta y}{\delta v(t)}$, тогда $Y(t,v) = y(v + a\chi(s,t))$ имеет производную $\frac{\partial Y(t,v)}{\partial t}$, причем

$$\frac{\partial Y(t,v)}{\partial t} = a(t)\frac{\delta y(v + a\chi(s,t))}{\delta v(t)}.$$

Теорема 2. Отображение

$$y = \varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2 - i\chi(0,t))\varphi_f(z)\Phi_1\left(\int_0^t a_{\varepsilon_1}(s)ds, x\right)$$

$$\times \Phi_1\left(i\int_{\mathbb{R}_+}\int_0^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x\right)\times\Phi_2\left(\frac{1}{2}\int_0^t\int_0^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)$$

$$*\left(M(u_0(x))\delta(t)\right)+\int_0^t \varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2-i\chi(\tau,t))$$

$$\times\varphi_f(z)\Phi_1\left(\int_\tau^t a_{\varepsilon_1}(s)ds,x\right)\Phi_1\left(i\int_{\mathbb{R}_+}\int_\tau^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x\right)$$

$$\times\Phi_2\left(\frac{1}{2}\int_\tau^t\int_\tau^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)*\left(-i\varphi_{\varepsilon_1}(v_1)\frac{\delta\varphi_f(z)}{\delta z(\tau,x)}\right)d\tau \qquad (7)$$

является решением задачи (5), (6).

Доказательство проведем подстановкой (7) в уравнения (5), (6). Частную производную по первому аргументу от функции g будем обозначать $D_1 g$. Используя предыдущую теорему 1, находим производную по переменной t

$$\frac{\partial y}{\partial t}=-i\varphi_{\varepsilon_1}(v_1)\frac{\delta\varphi_{\varepsilon_2}(v_2-i\chi(0,t))}{\delta v_2(t)}\varphi_f(z)\Phi_1\left(\int_0^t a_{\varepsilon_1}(s)ds,x\right)$$

$$\times\Phi_1\left(i\int_{\mathbb{R}_+}\int_0^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x\right)$$

$$\times\Phi_2\left(\frac{1}{2}\int_0^t\int_0^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)*\left(M(u_0(x))\delta(t)\right)$$

$$+\varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2-i\chi(0,t))\varphi_f(z)D_1\left(\Phi_1\left(\int_0^t a_{\varepsilon_1}(s)ds,x\right)\right)a_{\varepsilon_1}(t)$$

$$\times\Phi_1\left(i\int_{\mathbb{R}_+}\int_0^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x\right)$$

$$\times\Phi_2\left(\frac{1}{2}\int_0^t\int_0^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)*\left(M(u_0(x))\delta(t)\right)$$

$$+\varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2-i\chi(0,t))\varphi_f(z)\Phi_1\left(\int_0^t a_{\varepsilon_1}(s)ds,x\right)$$

$$\times D_1\left(\Phi_1\left(i\int_{\mathbb{R}_+}\int_0^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x\right)\right)\left(i\int_{\mathbb{R}_+}b_{\varepsilon_1}(s_1,t)v_1(s_1)ds_1\right)$$

$$\times\Phi_2\left(\frac{1}{2}\int_0^t\int_0^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)*\left(M(u_0(x))\delta(t)\right)$$

$$+\varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2-i\chi(0,t))\varphi_f(z)\Phi_1\left(\int_0^t a_{\varepsilon_1}(s)ds,x\right)$$

$$\times\Phi_1\left(i\int_{\mathbb{R}_+}\int_0^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x\right)$$

$$\times D_1\left(\Phi_2\left(\frac{1}{2}\int_0^t\int_0^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)\right)\int_0^t b_{\varepsilon_1}(s_1,t)\,ds_1 * \left(M(u_0(x))\delta(t)\right)$$

$$+\varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2)\varphi_f(z)\Phi_1^2(0,x)\Phi_2(0,x) * \left(-i\varphi_{\varepsilon_1}(v_1)\frac{\delta\varphi_f(z)}{\delta z(t,x)}\right)$$

$$-i\int_0^t \varphi_{\varepsilon_1}(v_1)\frac{\delta\varphi_{\varepsilon_2}(v_2-i\chi(\tau,t))}{\delta v_2(t)}\varphi_f(z)\Phi_1\left(\int_\tau^t a_{\varepsilon_1}(s)ds,x\right)$$

$$\times\Phi_1\left(i\int_{\mathbb{R}_+}\int_\tau^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x\right)$$

$$\times\Phi_2\left(\frac{1}{2}\int_\tau^t\int_\tau^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)d\tau * \left(-i\varphi_{\varepsilon_1}(v_1)\frac{\delta\varphi_f(z)}{\delta z(t,x)}\right)$$

$$+\int_0^t \varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2-i\chi(\tau,t))\varphi_f(z)D_1(\Phi_1\left(\int_\tau^t a_{\varepsilon_1}(s)ds,x\right))a_{\varepsilon_1}(t)$$

$$\times\Phi_1(i\int_{\mathbb{R}_+}\int_\tau^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x)$$

$$\times\Phi_2\left(\frac{1}{2}\int_\tau^t\int_\tau^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right) * \left(-i\varphi_{\varepsilon_1}(v_1)\frac{\delta\varphi_f(z)}{\delta z(\tau,x)}\right)d\tau$$

$$+\int_0^t \varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2-i\chi(\tau,t))\varphi_f(z)\Phi_1\left(\int_\tau^t a_{\varepsilon_1}(s)ds,x\right)$$

$$\times D_1(\Phi_1(i\int_{\mathbb{R}_+}\int_\tau^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x))$$

$$\times\left(i\int_{\mathbb{R}_+} b_{\varepsilon_1}(t,s_2)v_1(t)ds_2\right)\Phi_2\left(\frac{1}{2}\int_\tau^t\int_\tau^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)$$

$$*\left(-i\varphi_{\varepsilon_1}(v_1)\frac{\delta\varphi_f(z)}{\delta z(\tau,x)}\right)d\tau+\int_0^t \varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2-i\chi(\tau,t))\varphi_f(z)\Phi_1\left(\int_\tau^t a_{\varepsilon_1}(s)ds,x\right)$$

$$\times\Phi_1(i\int_{\mathbb{R}_+}\int_\tau^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x)$$

$$\times D_1(\Phi_2\left(\frac{1}{2}\int_\tau^t\int_\tau^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right))\int_\tau^t b_{\varepsilon_1}(s_1,t)ds_1 * \left(-i\varphi_{\varepsilon_1}(v_1)\frac{\delta\varphi_f(z)}{\delta z(\tau,x)}\right)d\tau \qquad (8)$$

Технику вариационного дифференцирования можно найти в [4]. Вычислим вариационную производную от y

$$\frac{\delta_p y}{\delta v_1(t)} = \varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2-i\chi(0,t))\varphi_f(z)\left(ia_{\varepsilon_1}(t)-\int_{\mathbb{R}_+} b_{\varepsilon_1}(s_1,t)v_1(s_1)ds_1\right)$$

$$\times\Phi_1\left(\int_0^t a_{\varepsilon_1}(s)ds,x\right)\Phi_1\left(i\int_{\mathbb{R}_+}\int_0^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x\right)$$

$$\Phi_2\left(\frac{1}{2}\int_0^t\int_0^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right) * \left(M(u_0(x))\delta(t)\right)$$

$$+\varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2-i\chi(0,t))\varphi_f(z)\Phi_1\left(\int_0^t a_{\varepsilon_1}(s)ds,x\right)$$

$$\times D_1(\Phi_1\left(i\int_{\mathbb{R}_+}\int_0^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x\right))\left(i\int_0^t b_{\varepsilon_1}(t,s_2)ds_2\right)$$
$$\times\Phi_2\left(\frac{1}{2}\int_0^t\int_0^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)*\left(M(u_0(x))\delta(t)\right)$$
$$+\int_0^t\varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2-i\chi(\tau,t))\varphi_f(z)\left(ia_{\varepsilon_1}(t)-\int_{\mathbb{R}_+}b_{\varepsilon_1}(s_1,t)v_1(s_1)ds_1\right)$$
$$\times\Phi_1\left(\int_\tau^t a_{\varepsilon_1}(s)ds,x\right)\times\Phi_1\left(i\int_{\mathbb{R}_+}\int_0^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x\right)$$
$$\times\Phi_2\left(\frac{1}{2}\int_\tau^t\int_\tau^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)*\left(-i\varphi_{\varepsilon_1}(v_1)\frac{\delta\varphi_f(z)}{\delta z(\tau,x)}\right)d\tau$$
$$+\int_0^t\varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2-i\chi(\tau,t))\varphi_f(z)\Phi_1\left(\int_\tau^t a_{\varepsilon_1}(s)ds,x\right)$$
$$\times D_1(\Phi_1\left(i\int_{R_+}\int_\tau^t b_{\varepsilon_1}(s_1,s_2)v_1(s_2)ds_1ds_2,x\right))$$
$$\times(i\int_0^t b_{\varepsilon_1}(t,s_2)ds_2\,\Phi_2\left(\frac{1}{2}\int_\tau^t\int_\tau^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)*\left(-i\varphi_{\varepsilon_1}(v_1)\frac{\delta\varphi_f(z)}{\delta z(\tau,x)}\right)d\tau$$
$$+\int_0^t\varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2-i\chi(\tau,t))\varphi_f(z)\Phi_1\left(\int_\tau^t a_{\varepsilon_1}(s)ds,x\right)$$
$$\times\Phi_1\left(i\int_{\mathbb{R}_+}\int_0^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x\right)\Phi_2\left(\frac{1}{2}\int_\tau^t\int_\tau^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)$$
$$*\left(-i\varphi_{\varepsilon_1}(v_1)\left(ia_{\varepsilon_1}(t)-\int_{\mathbb{R}_+}b_{\varepsilon_1}(s_1,t)v_1(s_1)ds_1\right)\frac{\delta\varphi_f(z)}{\delta z(\tau,x)}\right)d\tau. \qquad (9)$$

Подобным образом находятся другие вариационные производные из (5).

Поскольку Φ_1,Φ_2 – фундаментальные решения, то выполняются условия

$$D_1\Phi_1\left(\int_0^t a_{\varepsilon_1}(s)ds,x\right)=\frac{\partial^2}{\partial x^2}\Phi_1\left(\int_0^t a_{\varepsilon_1}(s)ds,x\right)+\delta\left(\int_0^t a_{\varepsilon_1}(s)ds,x\right),$$
$$D_1\Phi_2\left(\frac{1}{2}\int_0^t\int_0^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)$$
$$=\frac{\partial^4}{\partial x^4}\Phi_2\left(\frac{1}{2}\int_0^t\int_0^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)+\delta\left(\frac{1}{2}\int_0^t\int_0^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right),$$
$$\Phi_1(0,x)=\delta(x),\Phi_2(0,x)=\delta(x)$$

и оператор $\frac{\partial}{\partial x}$ коммутирует с Φ_1,Φ_2. Далее,

$$\delta\left(\int_0^t a_{\varepsilon_1}(s)ds,x\right)*\left(M(u_0(x))\delta(t)\right)=0,$$
$$M(u_0(x)\delta(t))*\delta\left(\frac{1}{2}\int_0^t\int_0^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)=0.$$

Подставляя найденные выражения в уравнение (5) и используя последние замечания, получаем тождество. Наконец,

$$y(0,x,v_1,v_2,z)=\varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2-i\chi(0,t))\varphi_f(z)$$
$$\times\Phi_1\left(\int_0^t a_{\varepsilon_1}(s)ds,x\right)\Phi_1\left(i\int_{\mathbb{R}_+}\int_0^t b_{\varepsilon_1}(s_1,s_2)v_1(s_1)ds_1ds_2,x\right)$$
$$\times\Phi_2\left(\frac{1}{2}\int_0^t\int_0^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2,x\right)*\left(M(u_0(x))\delta(t)\right)|_{t=0}$$

$$= \varphi_{\varepsilon_1}(v_1)\varphi_{\varepsilon_2}(v_2)\varphi_f(z)M(u_0(x)).$$

Следовательно, выполняется и начальное условие (6). Теорема доказана.

Математическое ожидание решения задачи (1), (2)

Теорема 3. Математическое ожидание решения задачи (1), (2) представимо в виде

$$M(u(t,x)) = exp\left(\int_0^t \varepsilon_2(\tau)d\tau + \frac{1}{2}\int_0^t\int_0^t b_{\varepsilon_2}(s_1,s_2)ds_1ds_2\right)\frac{\theta\left(\int_0^t a_{\varepsilon_1}(s)ds\right)}{\sqrt{2\pi\int_0^t a_{\varepsilon_1}(s)ds}}$$

$$\times\int_{\mathbb{R}} M(u_0(\xi))exp\left(-\frac{(x-\xi)^2}{4\int_0^t a_{\varepsilon_1}(s)ds}\right)\Phi_2\left(\frac{1}{2}\int_0^t\int_0^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2, x-\xi\right)d\xi$$

$$+\int_0^t\int_{\mathbb{R}} exp\left(\int_s^t \varepsilon_2(\tau)d\tau + \frac{1}{2}\int_s^t\int_s^t b_{\varepsilon_2}(s_1,s_2)ds_1ds_2\right)\frac{M(f(\tau,\xi))}{\sqrt{2\pi\int_\tau^t a_{\varepsilon_1}(s)ds}}$$

$$\times exp\left(-\frac{(x-\xi)^2}{4\int_\tau^t a_{\varepsilon_1}(s)ds}\right)\Phi_2\left(\frac{1}{2}\int_\tau^t\int_\tau^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2, x-\xi\right)d\xi d\tau. \quad (10)$$

Доказательство. Для нахождения математического ожидания $M(u(t,x))$ нужно в выражении (7) положить $v_1 = 0, v_2 = 0,\ z = 0$. При этом

$$\varphi_{\varepsilon_1}(0) = 1, \varphi_f(0) = 1, \Phi_1(0,x) = \delta(x), \Phi_2(0,x) = \delta(x),$$

$$\frac{\delta\varphi_f(z)}{\delta z(t,x)}\Big|_{z=0} = iMf(t,x), \varphi_{\varepsilon_2}(-i\chi(s,t)) = exp\left(\int_s^t \varepsilon_2(\tau)d\tau + \frac{1}{2}\int_s^t\int_s^t b_{\varepsilon_2}(s_1,s_2)ds_1ds_2\right).$$

В итоге получаем

$$M(u(t,x)) = \varphi_{\varepsilon_2}(-i\chi(0,t))\Phi_1\left(\int_0^t a_{\varepsilon_1}(s)ds, x\right)\Phi_2\left(\int_0^t\int_0^t b_{\varepsilon_1}(s_1,s_2)ds_1ds_2, x\right)\delta(x)$$

$$* M(u_0(x)\delta(t)) + \int_0^t \varphi_{\varepsilon_2}(-i\chi(\tau,t))\Phi_1\left(\int_\tau^t a_{\varepsilon_1}(s)ds, x\right)$$

$$\times\Phi_2\left(\int_\tau^t\int_\tau^t b_\varepsilon(s_1,s_2)ds_1ds_2, x\right)\delta(x) * M(f(t,x)).$$

Подставив выражение для Φ_1 и вычислив свертки, получаем (10). Теорема доказана.

Периодическое в среднем решение скалярного линейного уравнения теплопроводности со случайными коэффициентами. Рассмотрим частный случай уравнения (1), а именно уравнение теплопроводности вида

$$\frac{\partial y}{\partial t} = \frac{\partial^2 y}{\partial x^2} + \varepsilon(t)y + f(t,x), \qquad t \in \mathbb{R}_+, \qquad x \in \mathbb{R}, \quad (11)$$

при условии

$$y(0,x) = y_0(x),$$

где $\varepsilon(t)$ и $f(t,x)$ являются независимыми случайными процессами, заданными характеристическими функционалами $\varphi_\varepsilon(v)$ и $\varphi_f(z)$.

Приведем условия периодичности в среднем решения уравнения (11) при $t \in \mathbb{R}_+$ для гауссова случайного процесса $\varepsilon(t)$ с характеристическим функционалом вида

$$\varphi_\varepsilon(v) = exp\left(i\int_{\mathbb{R}_+} M(\varepsilon(s))v(s)ds - \frac{1}{2}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} b(s_1,s_2)v(s_1)v(s_2)ds_1ds_2\right), \quad (12)$$

где $M(\varepsilon(s))$ – математическое ожидание, а $b(s_1,s_2) = M(\varepsilon(s_1)\varepsilon(s_2)) - M(\varepsilon(s_1))M(\varepsilon(s_2))$ – ковариационная функция случайного процесса $\varepsilon(t)$.

Полугруппа, определяемая дифференциальным оператором второго порядка по переменной x, задается соотношениями

$$U_x(t)z(v,x) = \frac{1}{2\sqrt{\pi t}}\int_{\mathbb{R}} exp\left(-\frac{\tau^2}{4t}\right)z(v,x-\tau)d\tau, \qquad t \in \mathbb{R}_+,$$
$$U_x(0) = I,$$

где $v \in \mathbb{R}$ – некоторый параметр, I, как обычно, означает тождественный оператор.

Введем оператор $W(t,s)$ следующим образом:

$$W(t,s)\big(z(v,x)\varphi(u)\big) = U_x(t)U_x^{-1}(s)z(v,x)V(t,s)\varphi(u),$$

где оператор $V(t,s)\colon C(L_1(\mathbb{R}_+)) \to C(L_1(\mathbb{R}_+))$ определяется посредством соотношения $V(t,s)\varphi(u) = \varphi\big(u - i\chi(s,t)\big)$.

Теорема 4. Если $\varepsilon(t)$ – гауссов случайный процесс, $M\big(\varepsilon(t)\big)$ – ω-периодическая функция, $b(t,s)$ – ω-периодическая функция по обеим переменным, $M\big(f(t,x)\big)$ – ω-периодическая по $t \in \mathbb{R}_+$ непрерывная функция и оператор $I - W(\omega,0)$ обратим, тогда

$$M\big(y(t,x)\big) = \left[W(t,0)(W(0,\omega)-I)^{-1}\int_t^{t+\omega} W(0,s)(M(f(s,x))\varphi_\varepsilon(v))ds\right]|_{v=0}$$

является ω-периодическим по $t \in \mathbb{R}_+$ математическим ожиданием решения уравнения (11).

Следствие 1. Если случайный процесс $\varepsilon(t)$ задается характеристическим функционалом (12) и предположения теоремы 4 выполняются, тогда

$$M\big(y(t,x)\big) = [(W(0,\omega)-I)^{-1}\int_t^{t+\omega} U_x(t)U_x^{-1}(s)(M(f(s,x))exp(i\int_{\mathbb{R}_+} M\big(\varepsilon(s)\big)v(s)ds$$
$$-\int_t^s M\big(\varepsilon(s)\big)ds - \frac{1}{2}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} b\,(s_1,s_2)v(s_1)v(s_2)ds_1ds_2$$
$$-i\int_{\mathbb{R}_+}\int_t^s b\,(s_1,s_2)v(s_1)ds_2ds_1 + \frac{1}{2}\int_t^s\int_t^s b\,(s_1,s_2)ds_1ds_2)ds]\,|_{v=0}$$

является ω-периодическим по $t \in \mathbb{R}_+$ математическим ожиданием решения уравнения (11).

Замечание 1. Из обратимости оператора $I - W(\omega,0)$ следует, что однородное уравнение, соответствующее (11), не имеет отличных от нуля ω-периодических в среднем решений.

Замечание 2. Аналогичным образом периодические в среднем решения уравнения (11) могут быть получены для равномерно распределенного случайного процесса $\varepsilon(t)$ с характеристическим функционалом, заданным формулой

$$\varphi_\varepsilon(v) = \frac{\sin\int_{\mathbb{R}_+} a(s)v(s)ds}{\int_{\mathbb{R}_+} a(s)v(s)ds}\exp\left(i\int_{\mathbb{R}_+}\xi(s)v(s)ds\right),$$

где $a(s) \geq 0\ и\ \xi(s) = M(\varepsilon(s))$ – заданные непрерывные функции.

Если $a(s) \equiv 0$, тогда мы предполагаем, что $\varphi_\varepsilon(v) = exp\left(i\int_{\mathbb{R}_+}\xi(s)v(s)ds\right)$.

Заключение

Если рассматривать неоднородность в уравнении в качестве управляющей функции, то полученные результаты можно трактовать как нахождение управления, обеспечивающего периодичность математического ожидания решения.

Работа второго автора была поддержана Российским научным фондом (проект 17-11-01220).

ЛИТЕРАТУРА

1. Якубович В.А., Старжинский В.М. Линейные дифференциальные уравнения с периодическими коэффициентами и их приложения. М.: Наука, 1972. 720 с.
2. Задорожний В.Г., Курина Г.А. Периодические в среднем решения линейного дифференциального уравнения первого порядка. *Доклады академии наук.* 2013. Т. 450. № 5. С. 505-510.

3. Задорожний В.Г., Курина Г.А. Периодические в среднем решения линейного неоднородного дифференциального уравнения первого порядка со случайными коэффициентами. *Дифференциальные уравнения.* 2014. Т. 50. № 6. С. 726-744.
4. Задорожний В.Г. Методы вариационного анализа. М., Ижевск: РХД, 2006. 316 с.
5. Бутковский А.Г. Характеристики систем с распределенными параметрами. М.: Наука, 1979. 224 с.
6. Владимиров В.С. Обобщенные функции в математической физике. М.: Наука, 1976. 280 с.

V.G. Zadorozhniy (Voronezh State University, Voronezh), **G.A. Kurina** (Voronezh State University, Voronezh, Institute of Law and Economics, Voronezh, Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow)

On mean periodic solutions of linear inhomogeneous differential equations with random coefficients.

New results concerning the determination of the mathematical expectation for solutions of a linear diffusion equation with one spatial variable and random coefficients will be presented in the talk. For this we obtain an auxiliary non-random differential equation with ordinary and variational derivatives and a determinate initial condition.

For the particular case of the equation under consideration, namely the heat equation, in which the coefficient in front of an unknown function is Gaussian or uniformly distributed, and also for an ordinary inhomogeneous scalar linear first-order differential equation with random coefficients, conditions for the existence of mean periodic solutions are given.

If we consider the non-homogeneity as a control function, then obtained results can be interpreted as finding a control that ensures the periodicity of the mathematical expectation of a solution.

Относительный порядок дискретных моделей линейных стационарных систем

В.А.БОНДАРКО
Санкт-Петербургский Университет, математико-механический факультет

Рассмотрим функционирующую в непрерывном времени линейную стационарную систему с передаточной функцией $W(\lambda) = C(\lambda I - A) - 1B$, где I — единичная матрица, а A, B и C — матричные коэффициенты описывающего систему дифференциального уравнения в пространстве состояний:

$$\dot{x} = Ax + Bu,\ y = Cx.$$

Будем предполагать поначалу, что m, количество входов системы, совпадает с количеством выходов l.

Если управляющее воздействие (входной сигнал) постоянно на каждом из последовательных временных интервалов длиной $h > 0$, то равноотстоящие (на h) отсчеты входного $u_k = uhk$ и выходного $y_k = yhk$ сигналов связаны между собой разностным уравнением в пространстве состояний

$$x_{k+1} = Px_k + Qu_k,\ y_k = Cx_k$$

с коэффициентами $P = exp(hA)$, $Q = A^{-1}[exp(hA) - I]B, C$. Это уравнение описывает дискретную систему — результат дискретизации исходной системы по так называемому методу нулевого порядка (zero order hold), h называют шагом (или периодом) дискретизации. Название метода связано с тем, что на каждом из интервалов длиной h входной сигнал — константа, то есть многочлен нулевого порядка. Обозначим через $\chi(\lambda)$ передаточную функцию дискретной модели: $\chi(\lambda) = C(\lambda I - P)^{-1}Q$.

Если привести дискретную модель к форме «вход-выход», то в каждый момент дискретного времени самые последние значения входного сигнала будут входить в уравнение системы с коэффициентом $\beta = CQ$, который мы будем называть старшим коэффициентом дискретной системы-модели. Это и в самом деле старший коэффициент матричного числителя передаточной функции $\chi(\lambda)$, если ее знаменателем считать характеристический

многочлен матрицы P. Данная работа посвящена вопросу о том, как ведет себя $r(h)$ — ранг старшего коэффициента дискретной модели при стремлении к нулю шага дискретизации h.

Значение ранга β играет важную роль при решении многих задач теории управления. Так, например, в работах [1,2] А.Е.Барабанова по построению минимаксных регуляторов для дискретных систем одним из первых шагов служит искусственное приведение системы к такому виду, чтобы ранг β был максимален. Оказывается, если, как это часто бывает, дискретный объект управления получен в результате дискретизации невырожденной системы, то этот шаг— лишний, ибо условие о максимальном ранге автоматически выполняется при всех достаточно малых h.

Теорема. *Пусть* $l = m$ *и* $\det W(\lambda)$ *не равен нулю тождественно. Тогда* $r(h) = l$ *при всех достаточно малых* $h > 0$.

Следствие. *Для произвольных* l *и* m *при всех достаточно малых* h *значение* $r(h)$ *равно максимальной размерности минора* $W(\lambda)$*, не равного нулю тождественно.*

Утверждение, очевидно, остается верным при наличии транспортного запаздывания (одного и того же для всех компонент управления) и запаздывания в измерении (также одного для всех компонент выхода).

Обозначим i—тую строку произвольной матрицы T через $T[i]$. Определим R — векторный относительный порядок по Исидори [3], и интерактор Q:

$$R = \begin{bmatrix} R[1] \\ R[2] \\ \vdots \\ R[l] \end{bmatrix}, Q = \begin{bmatrix} C[1]A^{R[1]-1}B \\ C[2]A^{R[2]-1}B \\ \vdots \\ C[l]A^{R[l]-1}B \end{bmatrix},$$

где $R[i]$ — первое из натуральных чисел k, для которого $C[i]A^{k-1}B \neq 0$.

Из сформулированной теоремы следует, таким образом, что все компоненты векторного порядка дискретной модели при всех достаточно малых h имеют минимальное значение, то есть равны единице.

Лемма 1. *Пусть* $c \in \mathbb{R}^{m\times l}$ *и* $B = cAB = \ldots = cA^kB = 0$. *Тогда* $c(\lambda I - A)^{-1}A^{k+1}B = \lambda^{k+1}c(\lambda I - A)^{-1}B$

Доказательство банально следует из разложения резольвенты в ряд по степеням A.

Лемма 2. *Если*

$$\det C(\lambda I - A)^{-1}B \not\equiv 0, \tag{1}$$

то $\exists S:\ \det S = 1$ *и интерактор тройки* (A, B, SC) *не вырожден.*

Доказательство. Предположим, что, вопреки утверждению леммы, интерактор тройки (A, B, SC) вырожден для любой $(l \times l)$-матрицы S. Тогда, разумеется, и $\det Q = 0$. Положим $\nu = 0$ и $C_\nu = C_0 = C$.

Рассмотрим матрицу $\mathcal{C}_\nu$ со строками

$$\mathcal{C}_\nu[i] = C_\nu[i]A^{R_\nu[i]-1},$$

где $R_\nu[i]$ — относительный порядок тройки (A, B, C_ν) (при $\nu = 0$ он, естественно, совпадает с R). Относительный порядок тройки $(A, B, \mathcal{C}_\nu)$ состоит из одних единиц, но интерактор Q_ν у нее — тот же самый, что и у (A, B, C_ν). По лемме 1 передаточные функции этих троек связаны равенством

$$\mathcal{C}_\nu(\lambda I - A)^{-1}B = \begin{bmatrix} \lambda^{R_\nu[1]-1} & & & \\ & \lambda^{R_\nu[2]-1} & & \\ & & \ddots & \\ & & & \lambda^{R_\nu[l]-1}, \end{bmatrix} C_\nu(\lambda I - A)^{-1}B.$$

Следовательно,

$$\det \mathcal{C}_\nu(\lambda I - A)^{-1}B = \lambda^{\sum R[i]\,-l} \det C_\nu(\lambda I - A)^{-1}B. \tag{2}$$

Поскольку интерактор Q_ν — вырожденная матрица, то одна из ее строк — это линейная комбинация остальных:

$$\mathcal{C}_\nu[j]B = \sum_{i \neq j} s_i \mathcal{C}_\nu[i]B.$$

Положим

$$S = I - \begin{bmatrix} & & 0 & & \\ s_1 & \dots & s_{j-1} & 0 & s_{j+1} \; s_1 \; s_1 \\ & & 0 & & \end{bmatrix},$$

где во втором слагаемом отлична от нуля j-ая строка. Очевидно, $\det S = 1$. Обозначим $C_{\nu+1} = S\mathcal{C}_\nu$ и определим для тройки $(A, B, C_{\nu+1})$ ее векторный относительный порядок $R_{\nu+1}$ и интерактор $Q_{\nu+1}$. Интерактор останется вырожденным, поскольку $\det Q_{\nu+1} = \det S \det Q_\nu = 0$. Очевидно, далее, что $R_{\nu+1}[j] > 1$, поскольку

$$C_{\nu+1}[j]B = \mathcal{C}_\nu[j]B - \sum_{i \neq j} s_i \mathcal{C}_\nu[i]B = 0.$$

Остальные компоненты векторного порядка тройки $(A, B, C_{\nu+1})$ равны единице, поскольку i-ые строки матриц $\mathcal{C}_\nu$ и $C_{\nu+1}$ совпадают при $i \neq j$. Наконец,

$$\det C_{\nu+1}(\lambda I - A)^{-1}B = \det S \det \mathcal{C}_\nu(\lambda I - A)^{-1}B = \\ = \lambda^{\sum R_\nu[i]-l} \det C_\nu(\lambda I - A)^{-1}B. \qquad (3)$$

в силу (2).

Увеличим значение ν на единицу и повторим проделанные рассуждения. После nl шагов таких итераций мы получим

$$\det C_{nl+1}(\lambda I - A)^{-1}B = \lambda^K \det C(\lambda I - A)^{-1}B, \qquad (4)$$

где $K > nl$. Обозначим $a(\lambda) = \det(\lambda I - A)$. По детерминантной лемме Шура

$$\det C(\lambda I - A)^{-1}B = \frac{1}{a(\lambda)} \det \begin{bmatrix} \lambda I - A & B \\ -C & 0 \end{bmatrix}$$

Таким образом, тождество (4) эквивалено многочленному равенству

$$\det \begin{bmatrix} \lambda I - A & B \\ -C_{nl+1} & 0 \end{bmatrix} = \lambda^K \det \begin{bmatrix} \lambda I - A & B \\ -C & 0 \end{bmatrix}.$$

Оно может быть верно только в том случае, если нулю равны оба определителя в его левой и правой частях, а иначе степени левой и правой части будут различны. Следовательно, $\det C(\lambda I - A)^{-1}B \equiv 0 \; \forall \lambda$.

Итак, предположение о том, что для любой $(l \times l)$-матрицы S интерактор тройки (A, B, SC) вырожден, привело к противоречию с условием леммы — неравенством (1). Значит, это предположение ложно, то есть лемма доказана.

Доказательство теоремы. Требуется доказать невырожденность матрицы

$$\beta = CA^{-1}\left(e^{hA} - I\right)B = \\ = hC\left[I + \frac{1}{2}hA + \frac{1}{6}h^2A^2 + \ldots + \frac{1}{k!}h^{k-1}A^{k-1} + \ldots\right]B \qquad (5)$$

Лемма 2 гарантирует существование такой невырожденной матрицы S, что интерактор тройки (A, B, SC) невырожден. Не уменьшая общности, будем считать $S = I$, поскольку $\det(S\beta) = \det S \det \beta$.

Итак,

$$Q = \begin{bmatrix} C[1]A^{R[1]-1}B \\ C[2]A^{R[2]-1}B \\ \vdots \\ C[l]A^{R[l]-1}B \end{bmatrix},$$

— это невырожденный интерактор тройки (A, B, C), в котором $R[i]$ — первое из натуральных чисел k с ненулевым значением $C[i]A^{k-1}$. Ясно,что

$$h^{l-\sum R[i]} \det \beta_2 = \det[Q + \mathcal{O}(h)] \xrightarrow[h\to 0]{} \det Q \neq 0.$$

Теорема доказана.

Список литературы

[1] Барабанов А.Е., Граничин О.Н. Оптимальный регулятор линейного объекта с ограниченной помехой. Автоматика и телемеханика. 1984. № 5. С. 39-46.

[2] Барабанов А.Е. - СПб. : Изд-во С.-Петерб. ун-та, 1996. - 222 с.

[3] Isidori A. Nonlinear Control Systems. L.: Springer, 1995.

МНОЖЕСТВО ДОСТИЖИМОСТИ ДЛЯ МАШИНЫ ДУБИНСА

В. С. Пацко[1], **А. А. Федотов**[2]
(*Институт математики и механики им. Н.Н.Красовского УрО РАН,
Екатеринбург, 620990, ул. С.Ковалевской, 16,
patsko@imm.uran.ru*)

Аннотация

Представлено описание множества достижимости в момент для одной из самых популярных в задачах математической теории управления и в прикладных работах моделей управляемого движения на плоскости. Динамика движения задается с постоянной по величине линейной скоростью и оговоренным диапазоном возможных значений угловой скорости посредством нелинейной системы дифференциальных уравнений третьего порядка. Скалярное управление определяет текущую угловую скорость вращения вектора линейной скорости или, что эквивалентно, мгновенный радиус поворота. Допустимые значения управляющего параметра принадлежат замкнутому отрезку. Рассмотрены различные варианты расположения данного отрезка относительно нуля, включая случай одностороннего поворота (поворот возможен только в одну сторону). При классификации управлений, ведущих на границу множества достижимости, используется принцип максимума Понтрягина. Приведены результаты численных расчетов.

Ключевые слова: машина Дубинса, кусочно-постоянные управления, принцип максимума Понтрягина, трехмерное множество достижимости.

Введение

В прикладных работах, опирающихся на математическую теорию управления, очень популярной является модель управляемого объекта, называемая “машина Дубинса”. Такая модель задается нелинейной системой дифференциальных уравнений третьего порядка. Две фазовые переменные характеризуют геометрическое положение управляемого объекта на плоскости, третья переменная – угол направления вектора скорости. Величина скорости считается постоянной. Скалярное управляющее воздействие, стесненное геометрическим ограничением, определяет мгновенный радиус разворота.

Машина Дубинса – это управляемый объект (автомобиль или самолет) с простейшей моделью движения в горизонтальной плоскости. В 1957 г. американский математик Л. Дубинс опубликовал теоретическую работу [1] о линии кратчайшей длины с ограниченным радиусом кривизны, соединяющей две точки на плоскости с заданным направлением выхода из первой точки и заданным направлением входа во вторую. Полученные Л. Дубинсом результаты оказались очень полезными при исследовании объектов с ограниченным радиусом разворота и постоянной по величине скоростью передвижения. Именно поэтому такие объекты (особенно в работах по теоретической робототехнике, см., например, [2]) стали называть машиной Дубинса. В дальнейшем оказалось, что близкие вопросы в 1889 г. изучал А.А. Марков в работе [3], посвященной проблемам прокладки железных дорог. Динамика простейшего автомобиля использовалась Р. Айзексом в работах по дифференциальным играм [4, 5]. Модель Дубинса применяется при управлении колесными роботами, для диспетчерских расчетов в гражданской авиации, а также в прикладных работах по построению траекторий движения беспилотных летательных аппаратов в горизонтальной плоскости.

Под множеством достижимости $G(t_f)$ *в фиксированный момент времени* t_f понимаем совокупность всех состояний в трехмерном фазовом пространстве, в каждое из которых возможен перевод системы в момент t_f из заданного начального фазового состояния при помощи некоторого допустимого управления.

Традиционно при описании динамики машины Дубинса задаются ограничения снизу на радиусы разворота влево и вправо. Случай с одинаковыми ограничениями на радиус разворота влево и вправо будем называть симметричным, иначе – несимметричным. Несимметричный

[1] Кандидат физ.-мат. наук, ведущий научный сотрудник.
[2] Кандидат физ.-мат. наук, старший научный сотрудник.

случай возникает, например, при несбалансированности минимальных радиусов разворота влево и вправо. Множество достижимости в момент для симметричного и несимметричного случаев исследовалось ранее в работах [6 – 8]. Вообще говоря, задание ограничения на управление в виде отрезка может означать запрет на повороты влево или вправо, а также запрет на движение по прямой (если этот отрезок не содержит нулевую точку).

Предлагаемый доклад посвящен исследованию границы множества достижимости *в момент* для машины Дубинса в общем случае. Приведены примеры с результатами численных построений трехмерных множеств достижимости.

При исследовании границы множества достижимости в момент мы используем принцип максимума Понтрягина [9] для управлений, ведущих на границу множества достижимости [10]. Вся граница состоит из нескольких участков. Каждый участок характеризуется своим набором управлений, полностью определяемым двумя первыми переключениями, и, таким образом, является двухпараметрическим многообразием.

Из анализа структуры множества достижимости в момент вытекают некоторые полезные факты для задачи быстродействия (например, оценка числа переключений и характер переключений оптимального управления). Но в целом исследование границы множества достижимости в момент и решение задачи быстродействия (в частности, построение оптимального синтеза в задаче быстродействия) – это разные задачи. Для машины Дубинса известен синтез оптимального управления в задаче быстродействия для симметричного [11] и несимметричного [12] случаев.

Следует различать множества достижимости в момент и к моменту. Во втором случае момент перевода не является зафиксированным, а рассматривается на промежутке $[t_0, t_f]$. Для симметричного и несимметричного случаев построение множества достижимости к моменту рассматривалось в работе [7].

Постановка задачи

Пусть движение управляемого объекта (машина Дубинса) на плоскости описывается системой дифференциальных уравнений

$$\begin{aligned} \dot{x} &= \cos\varphi, \\ \dot{y} &= \sin\varphi, \\ \dot{\varphi} &= u, \quad u \in [u_1, u_2], \quad u_1 < u_2, \end{aligned} \tag{1}$$

где x, y – координаты геометрического положения, φ – угол наклона вектора скорости (рис. 1), u – управление, стесненное геометрическим ограничением. Величина линейной скорости равна 1. Далее будем считать, что $u_2 = 1$.

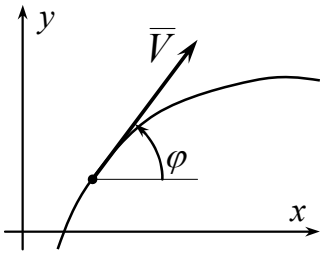


Рис. 1. Система координат

К виду (1) с $u_2 = 1$ путем нормировки геометрических координат и времени можно привести произвольную трехмерную систему, задающую движение с постоянной линейной скоростью и с управлением, которое определяет мгновенный радиус разворота. Положим в начальный момент времени $t_0 = 0$ начальное фазовое положение системы (1) равным $x_0 = 0$, $y_0 = 0$, $\varphi_0 = 0$. Допустимыми программными управлениями $u(\cdot)$ считаем измеримые функции времени со значениями в отрезке $[u_1, u_2]$. Угол φ рассматриваем на промежутке $(-\infty, +\infty)$.

Множество достижимости $G(t_f)$ в момент времени $t_f \geq t_0$ есть совокупность всех точек (x, y, φ) трехмерного фазового пространства, в каждую из которых возможен перевод системы (1) в момент t_f из начального положения (x_0, y_0, φ_0) при помощи некоторого допустимого управления на промежутке $[t_0, t_f]$.

В работе рассматривается построение множеств $G(t_f)$ для случаев $u_1 < 0$, $u_1 = 0$ и $u_1 > 0$.

Принцип максимума Понтрягина

Известно [10], что управления, которые ведут на границу множества достижимости, удовлетворяют принципу максимума Понтрягина (ПМП). Запишем соотношения принципа максимума для системы (1).

Пусть $u^*(\cdot)$ – некоторое допустимое управление, а $\left(x^*(\cdot), y^*(\cdot), x^*(\cdot)\right)^{\mathsf{T}}$ – соответствующее движение системы (1) на промежутке $[t_0, t_f]$. Дифференциальные уравнения сопряженной системы имеют вид

$$\begin{aligned} \dot{\psi}_1 &= 0, \\ \dot{\psi}_2 &= 0, \\ \dot{\psi}_3 &= \psi_1 \sin\varphi^*(t) - \psi_2 \cos\varphi^*(t). \end{aligned} \tag{2}$$

ПМП означает, что существует ненулевое решение $\left(\psi_1^*(\cdot), \psi_2^*(\cdot), \psi_3^*(\cdot)\right)^{\mathsf{T}}$ системы (2), для которого почти всюду (п.в.) на промежутке $[t_0, t_f]$ выполнено условие

$$\psi_1^*(t)\cos\varphi^*(t) + \psi_2^*(t)\sin\varphi^*(t) + \psi_3^*(t)\,u^*(t) = \max_{u\in[u_1,u_2]} \left(\psi_1^*(t)\cos\varphi^*(t) + \psi_2^*(t)\sin\varphi^*(t) + \psi_3^*(t)u\right).$$

Таким образом, условие максимума принимает вид

$$\text{(п.в.)}\quad \psi_3^*(t)\,u^*(t) = \max_{u\in[u_1,u_2]} \left(\psi_3^*(t)u\right), \qquad t\in[t_0,t_f]. \tag{3}$$

Отметим, что функции $\psi_1^*(\cdot)$ и $\psi_2^*(\cdot)$ есть константы. Обозначим их ψ_1^* и ψ_2^*. Если $\psi_1^* = 0$ и $\psi_2^* = 0$, то $\psi_3^*(\cdot) = \text{const} \neq 0$ на промежутке $[t_0, t_f]$. Следовательно, в этом случае либо п.в. $u^*(t) = u_1$, либо п.в. $u^*(t) = u_2$.

Пусть теперь хотя бы одно из чисел ψ_1^*, ψ_2^* не равно нулю. Опираясь на (1) и (2), можно записать следующее выражение для $\psi_3^*(t)$:

$$\psi_3^*(t) = \psi_1^* y^*(t) - \psi_2^* x^*(t) + C.$$

Отсюда следует, что $\psi_3^*(t) = 0$ тогда и только тогда, когда точка $\left(x^*(t), y^*(t)\right)^{\mathsf{T}}$ геометрического положения в момент t удовлетворяет уравнению прямой

$$\psi_1^* y - \psi_2^* x + C = 0. \tag{4}$$

Прямая (4) использовалась во многих работах, где для системы (1) анализировался принцип максимума.

В силу соотношения (3) имеем следующее. Если $\psi_3^*(t) > 0$ на некотором промежутке времени, то $u^*(t) = u_2$ п.в. на этом промежутке. Соответствующее движение в проекции на плоскость x, y идет в полуплоскости $\psi_1^* y - \psi_2^* x + C > 0$ по дуге окружности радиуса $1/u_2$ против часовой стрелки. Если $\psi_3^*(t) < 0$, то п.в. $u^*(t) = u_1$. Здесь движение идет в полуплоскости $\psi_1^* y - \psi_2^* x + C < 0$ по дуге окружности радиуса $1/u_1$ по часовой стрелке для $u_1 < 0$, по прямой для $u_1 = 0$ и по дуге окружности радиуса $1/u_1$ против часовой стрелки для $u_1 > 0$.

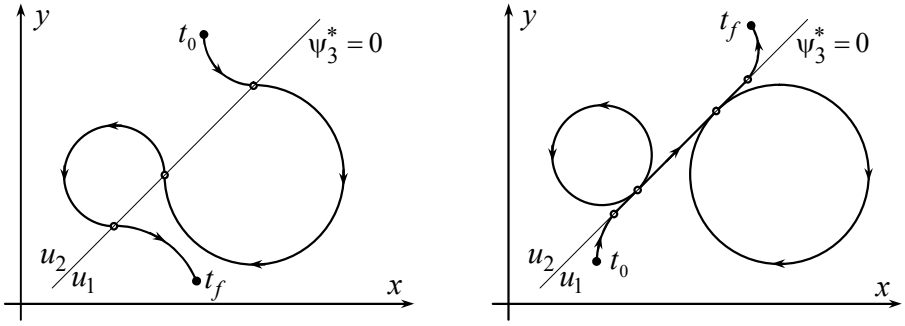
Условимся называть циклом участок движения длительностью 2π, на котором п.в. $u^*(t) = u_2$ или п.в. $u^*(t) = u_1 \neq 0$. Траектория движения на таком участке в проекции на плоскость x, y представляет собой окружность.

Если $\psi_3^*(t) = 0$ п.в. на некотором промежутке времени, то на этом промежутке движение $\left(x^*(t), y^*(t)\right)^{\mathsf{T}}$ идет по прямой (4). Стало быть, $\varphi^*(t) = \text{const}$. Поэтому $u^*(t) = 0$ п.в. на этом промежутке. Данный случай невозможен при $u_1 > 0$.

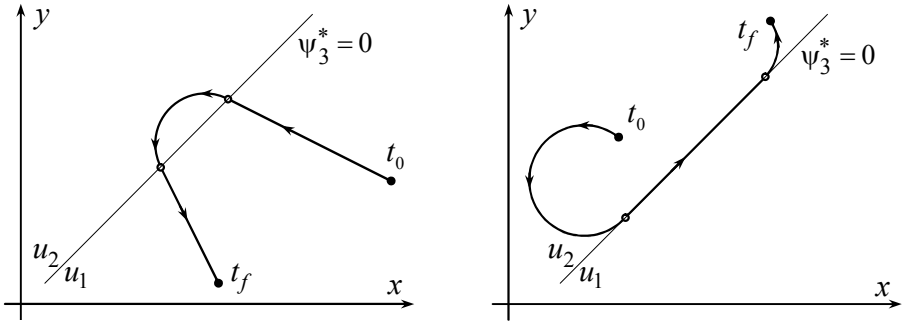
Очевидно, что управления, ведущие на границу множества достижимости и, соответственно, удовлетворяющие ПМП, могут изменять свое значение только на прямой (4), которую будем называть прямой переключения.

Рассмотрев варианты возможного взаимного расположения траектории движения $(x^*(t), y^*(t))^\mathsf{T}$ и прямой переключения (рис. 2), сформулируем следующее утверждение.

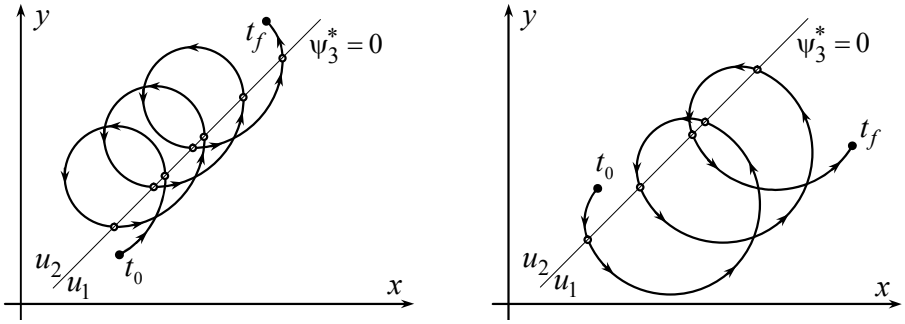
Утверждение 1. Пусть движение системы (1) удовлетворяет ПМП. Тогда соответствующая траектория состоит из конечного числа дуг окружностей и прямолинейных участков. Последнее возможно лишь при $0 \in [u_1, u_2]$.



а) Траектории $(x^*(t), y^*(t))^\mathsf{T}$ для случая $u_1 < 0$ (пример с $u_1 = -0.5$)



б) Траектории $(x^*(t), y^*(t))^\mathsf{T}$ для случая $u_1 = 0$



в) Траектории $(x^*(t), y^*(t))^\mathsf{T}$ для случая $u_1 > 0$ (пример с $u_1 = +0.5$)

Рис. 2. Траектории принципа максимума и прямая переключения

Если условие (3) выполнено, то в силу Утверждения 1 функция $\psi_3^*(\cdot)$ на промежутке $[t_0, t_f]$ может изменять свой знак только конечное число раз. Поэтому, мы можем считать управление $u^*(\cdot)$ (ведущее движение на границу множества достижимости $G(t_f)$) кусочно-постоянным с конечным числом переключений на $[t_0, t_f]$. Для определенности условимся считать такое управление кусочно-непрерывным справа и момент t_f не будем включать в совокупность моментов переключения.

Случай $u_1 < 0$

В работах [6 – 8] симметричный (с ограничениями $u_1 = -1$, $u_2 = +1$) и несимметричный ($u_1 < 0$, $u_2 = +1$) случаи были исследованы. Были доказаны теоремы о количестве и характере переключений управлений, ведущих на границу множества достижимости. А именно, было показано, что в любую точку границы множества достижимости $G(t_f)$ можно попасть при помощи управления, имеющего не более двух переключений. При этом можно ограничиться шестью вариантами последовательности управлений:

$$\begin{gathered} 1)\ u_2, 0, u_2\,;\ 2)\ u_1, 0, u_2\,;\ 3)\ u_2, 0, u_1\,; \\ 4)\ u_1, 0, u_1\,;\ 5)\ u_2, u_1, u_2\,;\ 6)\ u_1, u_2, u_1\,. \end{gathered} \qquad (5)$$

Для построения границы множества $G(t_f)$ перебираем все управления вида 1 – 6 из списка (5) с двумя моментами переключения t_1, t_2. Параметр t_1 выбирается из промежутка $[0, t_f]$, а параметр t_2 – из промежутка $[t_1, t_f]$. Управления с одним переключением и без переключений при этом также охватываются. Взяв конкретный вариант переключений и перебирая для него параметры t_1, t_2 на некоторой достаточно мелкой сетке, получаем набор точек, образующих поверхность в трехмерном пространстве x, y, φ. Каждому из шести вариантов в списке (5) соответствует своя поверхность в трехмерном пространстве. Граница множества достижимости $G(t_f)$ составляется из кусков этих поверхностей.

Рис. 3 показывает с двух ракурсов границу множества $G(t_f)$ для момента $t_f = 2.5\pi$. Различные участки границы выделены своим цветом. С шагом $(15/\pi)^\circ$ по оси φ изображены сечения множества достижимости плоскостями $\varphi = \text{const}$.

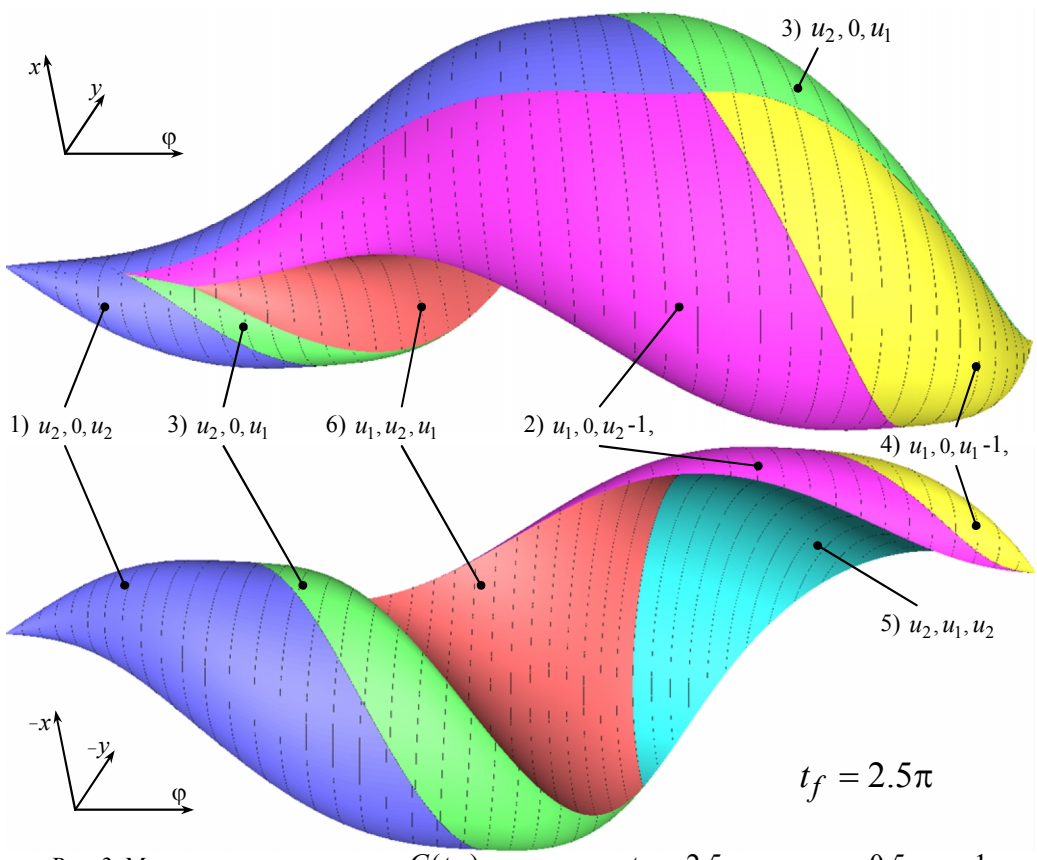


Рис. 3. Множество достижимости $G(t_f)$ для момента $t_f = 2.5\pi$ при $u_1 = -0.5, u_2 = 1$

Управление тождественно равное нулю ведет в точку стыковки участков 1 – 4. В точки линий, лежащих на стыке участков 1 – 2, 1 – 3, 2 – 4, 2 – 5, 2 – 6, 3 – 4, 3 – 5 и 3 – 6 ведет управление с одним переключением. В любую точку линии, являющейся общей для участков 5 и 6, идут два движения, каждое с двумя переключениями. На этой линии участки 5, 6 имеют негладкую стыковку.

На рис. 4 изображены в одном ракурсе множества достижимости $G(t_f)$ для четырех моментов времени t_f. Четко прослеживается изменение структуры границы множества достижимости. С увеличением времени передняя часть границы, составленная из участков 1 – 4, «затягивает» тыльную часть, составленную из участков 5, 6.

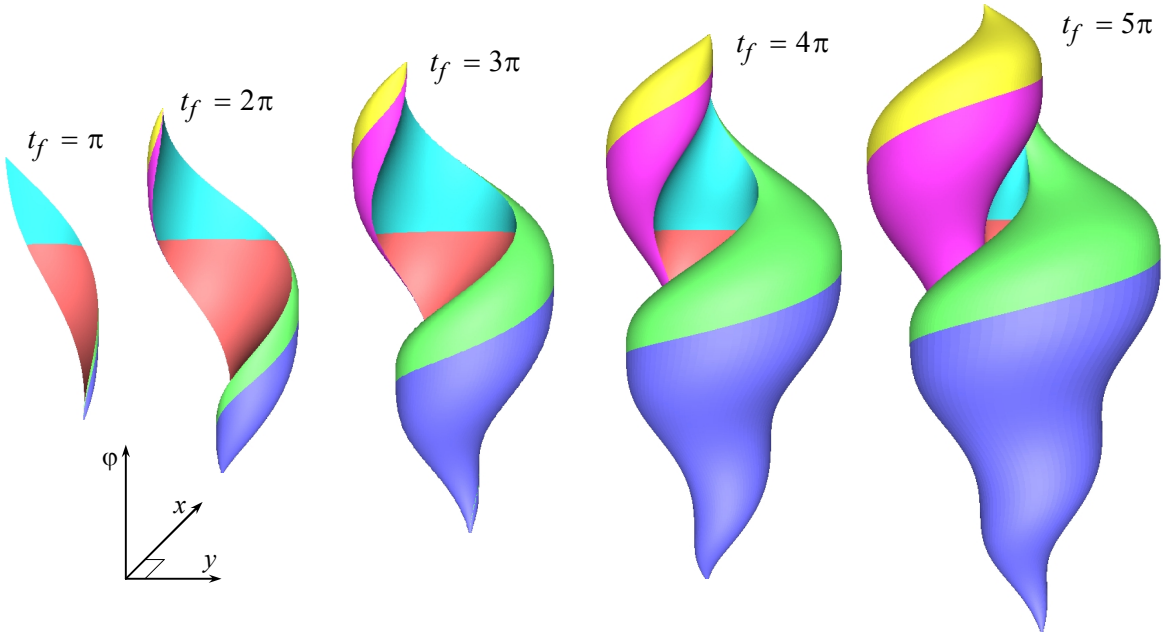


Рис. 4. Развитие во времени множеств достижимости $G(t_f)$ для $u_1 = -0.5,\ u_2 = 1$

Случай $u_1 = 0$

В этом случае, управления, ведущие на границу множества достижимости, принимают крайние значения $u_1 = 0$, $u_2 = 1$. Главная особенность в том, что прямолинейные участки могут появиться в двух ситуациях:

1) Когда соответствующий участок лежит в полуплоскости $\psi_1^* y - \psi_2^* x + C < 0$, где $\psi_3^*(t) < 0$. В этом случае максимум в соотношении (3) реализуется при $u(t) = 0$.

2) Когда система движется по линии переключения $\psi_1^* y - \psi_2^* x + C = 0$.

Если для движения, которое приводит на границу множества $G(t_f)$, выполнено $\psi_3^*(t_0) < 0$ для начального момента $t_0 = 0$, тогда либо движение является прямолинейным (в направлении начального вектора скорости) на всем промежутке $[t_0, t_f]$, либо оно достигает прямую переключения с некоторым ненулевым углом в некоторый момент и дальше идет по дуге окружности в полуплоскости $\psi_3^*(t) > 0$. Если до момента t_f траектория снова достигает прямую переключения, то вход происходит под тем же самым углом и траектория переходит в полуплоскость $\psi_3^*(t_0) < 0$ и далее является прямолинейной. Поэтому в случае $\psi_3^*(t_0) < 0$ траектория имеет не более двух переключений.

Пусть $\psi_3^*(t_0) > 0$. Тогда на начальном промежутке времени движение идет по дуге окружности. Когда оно достигает прямую переключения с некоторым ненулевым углом, то происходит

переход в полуплоскость $\psi_3^*(t_0) < 0$, где движение продолжается как прямолинейное. Если первое попадание на прямую переключения происходит по касательной, то в дальнейшем движение идет либо по дуге окружности, либо по линии переключения с возможным переходом в полуплоскость $\psi_3^*(t) > 0$. Если момент t_f является достаточно большим, тогда возможны циклы, которые были отмечены в разделе, посвященном принципу максимума. Такие циклы могут быть перенесены в начальную или, наоборот, в конечную часть движения. С учетом этого можно считать, в любую точку на границе множества достижимости можно попасть при помощи не более двух переключений.

Аналогично анализируется случай $\psi_3^*(t_0) = 0$.

Таким образом, справедливо

Утверждение 2. Пусть $u_1 = 0$. Тогда в любую точку на границе множества достижимости $G(t_f)$ ведет движение не более чем с двумя переключениями. При этом возможны только два варианта последовательности управлений: u_1, u_2, u_1; $. u_2, u_1, u_2$. Для первого варианта продолжительность второй части траектории (где $u(t) = u_2 = 1$) меньше, чем 2π.

В этом утверждении варианты только с одним переключением и без переключений также рассматриваются. Для этого длины одного или двух интервалов полагаются равными нулю.

В силу указанного свойства (по аналогии с построениями для случая $u_1 < 0$), граница множества достижимости $G(t_f)$ может быть построена при помощи двух поверхностей, соответствующих двум вариантам управления, описанным выше. На рис. 6 показаны примеры множества $G(t_f)$ для моментов $t_f = 4\pi, 6\pi$.

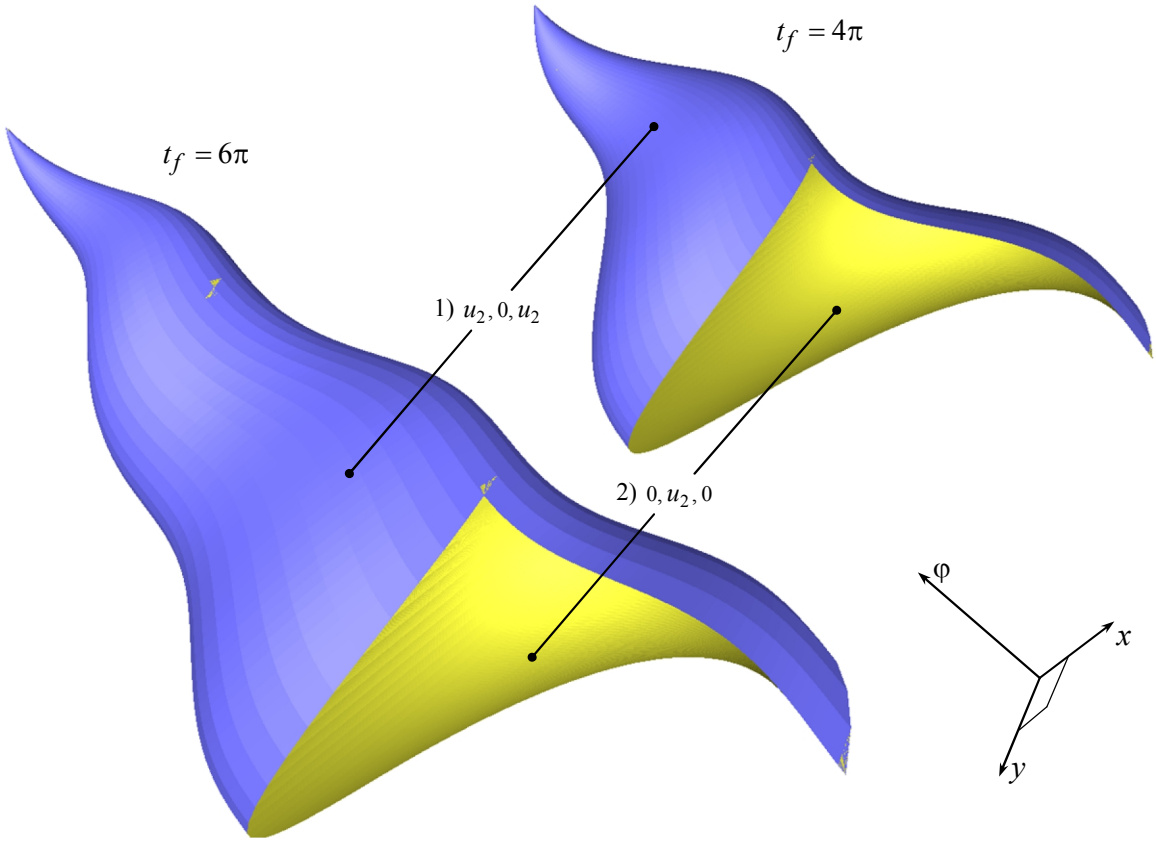


Рис.6. Множество достижимости $G(t_f)$ при $u_1 = 0$ для моментов $t_f = 4\pi, 6\pi$

В работе [14] установлено, что сечения множества $G(t_f)$ по угловой координате $\varphi = \text{const}$ в случае $u_1 = 0$ являются выпуклыми и имеют форму круга или кругового сегмента.

Случай $u_1 > 0$

В последнее время появляются работы (см., например, [13]), посвященные прикладным задачам, в которых возникает ситуация одностороннего поворота с исключением движения по прямой. В этом случае управления, ведущие на границу множества достижимости, представляют собой набор дуг окружностей радиусом $1/u_1$ и $1/u_2$. Число переключений является конечным [14], но растет с ростом момента t_f, на который строится множество достижимости.

Рассмотрим два соседних участка движения, удовлетворяющего принципу максимума, каждый из которых не является крайним на промежутке $[t_0, t_f]$. Суммарное изменение угла φ на объединении (склейке) двух таких участков равно 2π. Время прохождения указанной склейки на каждом движении является постоянным и находится в интервале $[2\pi/u_2, 2\pi/u_1]$.

При численном построении множества достижимости задается некоторое разбиение интервала возможных значений угла φ при заданном значении момента t_f. При фиксированном значении φ для построения соответствующего сечения множества $G(t_f)$ используются четыре типа управлений: 1) $u_2, ..., u_1$; 2) $u_1, ..., u_2$; 3) $u_2, ..., u_1, u_2$; 4) $u_1, ..., u_2, u_1$. Многоточие означает последовательность чередующихся управлений u_1 и u_2. Для случая 1 (соответственно, 2) повторяющиеся управления имеют вид набора пар управлений u_1, u_2 (u_2, u_1). Продолжительность действия каждой пары одна и та же для конкретного движения и находится в интервале $[2\pi/u_2, 2\pi/u_1]$. Количество таких пар может быть различным при $t_f > 2\pi$. Для случая 3 (соответственно, 4) повторяющиеся управления представляют собой пары u_1, u_2 (u_2, u_1) с дополнительным участком u_1 (u_2) перед последним участком, где действует управление u_2 (u_1). Длительности указанных участков определяются соотношениями, вытекающими из детального и нетривиального анализа ПМП.

На рис. 7 показан пример расчета границы множества достижимости с раскраской четырьмя цветами, соответствующими четырем указанным типам управлений. Отметим, что отдельные φ – сечения (за исключением крайних) множества $G(t_f)$ формируются управлениями от двух до четырех типов.

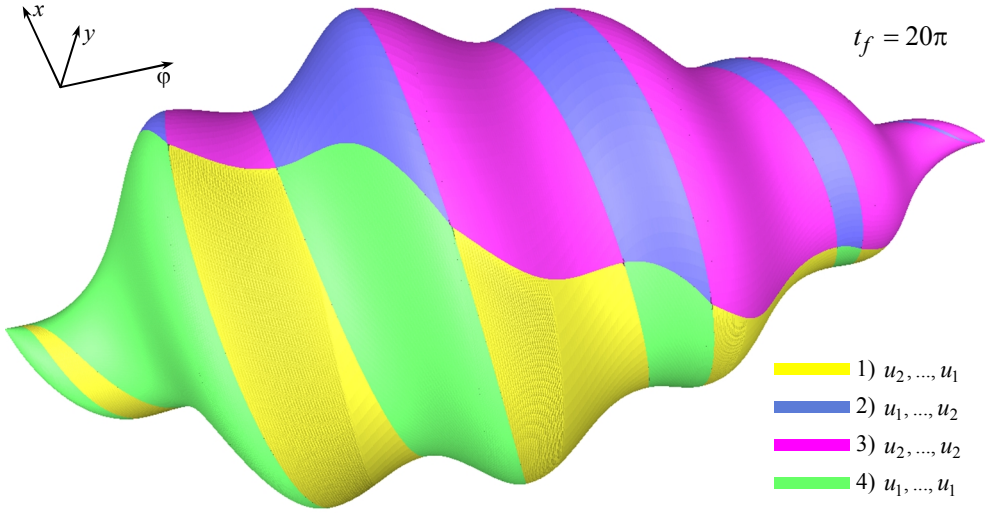


Рис. 7. Множество достижимости $G(t_f)$ при $u_1 = 0.5, u_2 = 1$ для момента $t_f = 20\pi$.

В процессе численных построений множества достижимости была выявлена выпуклость его φ – сечений. Краткое доказательство такого факта приведено в работе [15].

Замечание. Как было отмечено в постановке задачи, мы считаем возможные изменения угла φ в промежутке $(-\infty, +\infty)$. В инженерной практике обычно угол φ рассматривается в проме-

жутке размером 2π. Соответствующее множество достижимости легко строится на основе рассмотренного в работе подхода путем «нарезки» по угловой координате полученного множества достижимости на участки длиной 2π. Некоторые примеры приведены в работе [7].

Заключение

Решена задача построения множеств достижимости в момент для управляемого объекта «машина Дубинса» при задании ограничений на управление общего вида, включая случай одностороннего поворота. Представлены результаты численного моделирования. Дальнейшие исследования будут посвящены детальному описанию сечений множества достижимости по угловой координате.

Работа проводилась при поддержке гранта РФФИ 18-01-00410

ЛИТЕРАТУРА

1. Dubins L.E. On curves of minimal length with a constraint on average curvature and with prescribed initial and terminal positions and tangents // American Journal of Mathematics. 1957. Vol. 79, No. 3. P. 497–516.
2. Laumond J.-P. (ed.) Robot Motion Planning and Control, Lecture Notes in Control and Information Sciences, Vol. 229 / Springer-Verlag, Berlin Heidelberg, 1998. 354 p.
3. Марков А.А. Несколько примеров решения особого рода задач о наибольших и наименьших величинах // Сообщ. Харьков. матем. общ. 1889. 2-я сер., том 1, выпуск 2. С. 250–276.
4. Isaacs R. Games of pursuit / Scientific report of the RAND Corporation, Santa Monica, 1951.
5. Айзекс Р. Дифференциальные игры. М.: Мир, 1967. 479 с.
6. Пацко В.С., Пятко С.Г., Федотов А.А. Трехмерное множество достижимости нелинейной управляемой системы // Известия РАН. ТиСУ. 2003. № 3. С. 8–16.
7. Fedotov A., Patsko V., Turova V. Reachable sets for simple models of car motion / Recent Advances in Mobile Robotics, edited by Andon Venelinov Topalov, Rijeka, Croatia: InTech, 2011. P. 147–172. URL: http://home.imm.uran.ru/kumkov/Intech_paper_2011/Intech_paper.pdf.
8. Симоненко А.С., Федотов А.А. Множество достижимости для автомобиля Дубинса при несимметричном ограничении на управление [Электронный ресурс] / CEUR-WS, Proceedings of 48th International Youth School-conference: Modern Problems in Mathematics and its Applications, Yekaterinburg, Russia, 2017. Vol. 1894. P. 79–87, URL: http://ceur-ws.org/Vol-1894/opt6.pdf.
9. Понтрягин Л.С., Болтянский В.Г., Гамкрелидзе Р.В., Мищенко Е.Ф. Математическая теория оптимальных процессов. М.: Наука, 1969. 384 с.
10. Ли Э.Б., Маркус Л. Основы теории оптимального управления. М.: Наука, 1972. 576 с.
11. Pecsvaradi T. Optimal horizontal guidance law for aircraft in the terminal area // IEEE Trans. on Automatic Control. 1972. Vol. 17, No. 6. P. 763–772.
12. Bakolas E., Tsiotras P. Optimal synthesis of the asymmetric sinistral/dextral Markov-Dubins problem // Journal of Optimization Theory and Applications. 2011. Vol. 150, No. 2. P. 233–250.
13. Choi H. Time-Optimal Paths for a Dubins Car and Dubins Airplane with a Unidirectional Turning Constraint: Dissertation for the degree of Doctor of Philosophy / University of Michigan, 2014. 134 p.
14. Пацко В.С., Федотов А.А. Множество достижимости в момент для машины Дубинса в случае одностороннего поворота // Труды института математики и механики, 2018, № 1, Том 24, С. 143–155.
15. Patsko V.S., Fedotov A.A. Attainability set at instant for one-side turning Dubins car // Proceedings of 17th IFAC Workshop on Control Applications of Optimization, Yekaterinburg, Russia, 2018 (принято к печати).

V.S. Patsko, A.A. Fedotov (Krasovskii Institute of Mathematics and Mechanics, Yekaterinburg)

Reachable set for the Dubins car

The paper presents a description of reachable set at instant for a model, which is very popular in problems of mathematical control theory and applied problems. The motion dynamics is defined by a third order non-linear system of differential equation having the linear velocity of a constant magnitude and a prescribed range of possible values of the angular velocity. The scalar control is the instantaneous angular velocity of the linear velocity vector rotation or, which is equivalent, the instantaneous turn radius. Admissible control values belong to a closed interval. Different mutual locations of this interval and the origin are considered including the situation of a one-side turn. During classifying controls guiding the system to the boundary of the reachable set, we use the Pontryagin's maximum principle. Results of numerical computations are given.

РАЗРАБОТКА И ОЦЕНКА ЭФФЕКТИВНОСТИ МАЛОПОТРЕБЛЯЮЩИХ ПОЛУПРОВОДНИКОВЫХ СХЕМ ДЛЯ ЦИФРОВОЙ ОБРАБОТКИ СИГНАЛОВ НА ОСНОВЕ АНТИМАШИН ХАРТЕНШТЕЙНА

С. И. Салищев[1]

(СПбГУ, Россия, 199034, С.-Петербург, Университетская набережная 7-9, s.salischev@spbu.ru)

Аннотация

Рассматривается подход к разработке малопотребляющих полупроводниковых логических схем ориентированных на выполнение эффективных параллельных вычислений, характерных для алгоритмов цифровой обработки сигналов на основе непроцессорных архитектур, состоящих из крупных многофункциональных вычислительных блоков, связанных реконфигурируемым графом соединений, также называемых антимашинами Хартенштейна. Предлагается модель энергопотребления таких устройств Решается задача выбора оптимального по энергопотреблению параллелизма для устройства.

Ключевые слова: цифровая обработка сигналов, логические схемы, реконфигурируемые вычислительные системы, параллельные вычисления, энергоэффективные вычисления.

Введение

Энергоэффективность является одним из основных параметров беспроводных полупроводниковых устройств. Оптимизация энергоэффективности возможна на различных уровнях: транзисторном, вентильном, архитектурном и системном. Вопросы оптимизации на вентильном и транзисторном уровне многократно рассмотрены и лежат в области физики полупроводников. Алгоритмические оптимизации, относящиеся к системному уровню, как правило, рассматриваются в рамках существующих микропроцессорных архитектур и направлены на повышение производительности, а не энергоэффективности.

Архитектурные оптимизации тесно связаны с алгоритмическими, поскольку определяют относительную стоимость операций, модель памяти и параллелизма. Однако обычно они не рассматриваются совместно с алгоритмическими оптимизациями для улучшения энергопотребления, поскольку для фиксированной процессорной архитектуры они заранее выбраны и являются ограничениями, а не параметрами оптимизации.

Одним из стандартных подходов к повышению энергоэффективности является разработка специализированных полупроводниковых логических схем с фиксированной функциональностью для реализации определенного набора алгоритмов. Однако этот подход обладает несколькими серьезными недостатками. Во-первых, он ограничивает гибкость полученной системы, поскольку переиспользование ее для реализации других алгоритмов становится невозможным. Алгоритмы в области обработки сигналов и анализа данных постоянно развиваются, а время разработки и изготовления полупроводниковой схемы составляет не менее 6 месяцев. Таким образом, схема может устареть уже в процессе подготовки к производству.

Ошибки алгоритма, оказавшиеся в аппаратной реализации устройства, невозможно исправить без выпуска новой версии кристалла, что означает высокую стоимость ошибки. Это определяет консервативность в выборе алгоритмов и дополнительные усилия на проверку правильности их работы в процессе разработки аппаратной части устройства, что еще более замедляет процесс разработки. Кроме того, схему невозможно использовать в других аналогичных устройствах, имеющих немного отличающуюся функциональность (например, сотовый телефон, беспроводная гарнитура, наголовное устройство, часы), если эта вариативность не была заложена при разработке. Во-вторых, в чипах, произведенных по современным литографическим процессам, существенную долю потерь энергии создают токи утечки, которые пропорциональны суммарной площади подключенных к питанию элементов на кристалле. Таким образом, для минимизации токов утечки крупные базовые блоки, такие, как умножители, деревья сумматоров, управляемые сдвиговые регистры, регистровая память, статическая память долж-

[1] Кандидат физико-математических наук, старший преподаватель кафедры информатики.

ны максимально переиспользоваться между различными частями алгоритмов и, по возможности, полностью отключаться от питания во время простоя.

Это требование идеологически близко к концепции программируемого процессора. Под программируемостью подразумевается наличие эффективных инструментальных средств разработки программ на языках высокого уровня (C, C++, Fortran и др.) и возможность эффективной компиляции программ, реализующих широкий класс алгоритмов цифровой обработки сигналов.

Разработка на основе программируемого процессора обладает следующими преимуществами перед специализированным непрограммируемым устройством:

- возможность переиспользования вычислительных ресурсов между частями алгоритма для экономии площади кристалла и улучшения энергоэффективности;
- переиспользование памяти данных различными частями алгоритмов для уменьшения ее размера и уменьшения необходимости копирования данных, что ведет к улучшению энергоэффективности;
- возможность исправления ошибок в алгоритме после изготовления кристалла, что позволяет разделить процессы разработки аппаратуры и программы для сокращения общего времени разработки;
- возможность переиспользования блока и архитектуры кристалла в других типах устройств за счет изменения алгоритма для сокращения времени разработки и стоимости производства устройства;
- наличие развитых средств программирования (компилятор, отладчик, операционная система, библиотеки) для сокращения времени разработки.

С другой стороны, сигнальные процессоры общего применения во многих случаях оказываются недостаточно производительны и/или энергоэффективны для использования в малопотребляющих автономных устройствах по следующим причинам:

- дополнительные блоки для реализации программируемости;
- недостаточный параллелизм, связанный с последовательной структурой программы;
- избыточная ширина стандартных типов данных, требующая более широкой памяти и
- более сложных вычислительных блоков;
- использование части полосы пропускания памяти для загрузки программы;
- потери времени на выполнение ветвлений и циклов;
- малая гранулярность универсальных операций, приводящая к избыточным доступам в память.

Таким образом, разработка специализированных энергоэффективных систем цифровой обработки данных с высокой производительностью и возможностью переиспользования ресурсов представляется наиболее выигрышной стратегией как с точки зрения гибкости, так и с точки зрения энергоэффективности. Общая архитектура для таких вычислительных блоков на базе ОЗУ была предложена Хартенштейном [1] и основана на потоке данных, в отличие от обычных процессоров, управляемых программой (поток управления). Архитектура управляется с помощью шаблонов адресного генератора и конфигураций потока данных, таким образом, она может реализовывать произвольный наперед заданный параллелизм и переиспользование ресурсов без накладных расходов на исполнение программы.

Эти вычислительные блоки затем могут быть интегрированы с расширяемым процессорным ядром как реализация расширения набора инструкций или как акселераторы вычисления библиотечных функций. Таким образом, сохраняется и гибкость программируемого процессора и высокая производительность и энергоэффективность специализированной полупроводниковой схемы. Эта стратегия была взята за основу в данной работе.

Поскольку разработка эффективных инструментальных средств программирования является крайне сложной и ресурсоемкой задачей, то процессор должен основываться на стандартной микроархитектуре и системе команд с набором расширений, ускоряющих выполнение алгоритмов, специфичных для данной области. Такими стандартными расширяемыми микроархитектурами являются ARM, ARC, MIPS, Sparc. Все они имеют отличную поддержку такими инструментальными средствами разработки с открытым исходным кодом, как GNU GCC и

Clang/LLVM. Открытая лицензия на средства разработки позволяет сравнительно легко адаптировать их под требуемые наборы расширений.

В рамках специализации процессора решаются две задачи. Во-первых, выбирается архитектура памяти и представление данных, наиболее удобное для данного класса алгоритмов, например, ширина вектора для векторных вычислений, ширина слова, использование чисел с плавающей точкой, наличие штрафа за невыровненный доступ к памяти, ширина и количество дополнительных векторных регистров.

Во-вторых, выбирается набор дополнительных инструкций процессора, ускоряющих выполнение программ, характерный для данного класса алгоритмов, например, сдвиг с округлением, сложение с насыщением, нахождение старшего ненулевого разряда, КИХ фильтрация, бабочка БПФ, вычисление элементарных функций, битовые перестановки и т.д.

Задача выбора архитектуры памяти является наиболее сложной и важной, поскольку большинство алгоритмов при распараллеливании и специализации оказываются ограничены именно пропускной способностью памяти, а не вычислительными блоками. Кроме того, архитектура памяти определяет контекст реализации алгоритма и может существенно влиять на ход вычислений. При этом предпочтительным является случай, когда процессорное ядро и специализированный акселератор имеют общую память, поскольку уменьшаются накладные расходы на копирование данных.

Отличительной особенностью программируемых процессоров является наличие программы в памяти, что увеличивает необходимую пропускную способность памяти для выборки инструкций. Обычно специализированные процессоры имеют Гарвардскую архитектуру, которая в отличие от традиционной для процессоров общего применения архитектуры фон Неймана имеет дополнительную память программ, соединенную с процессором отдельной шиной. Это устраняет так называемое «узкое горлышко фон Неймана», когда доступы к программе и данным конкурируют за шину памяти. Следующим узким местом становится процесс декодирования и исполнения команд. Для упрощения блока декодирования инструкций большинство специализированных процессоров декодируют инструкции по одной в порядке выборки из памяти. Инструкции декодируются в темпе одной инструкции за такт синхронизации. Современные процессоры выполняют инструкции в конвейере. То есть одна инструкция проходит несколько стадий обработки: выборка из памяти, декодирование, чтение данных из регистров или памяти, вычисление результата, вычисление адреса следующей инструкции, запись результата в регистры или память. Если стадия выполнения инструкции занимает несколько тактов синхронизации, то результаты последовательных инструкций могут переупорядочиваться, в случае, если они независимы по данным.

Большинство алгоритмов, используемых в обработке сигналов и адаптивном управлении, состоят в применении небольшого вычислительного ядра к значительному объему данных. В программном коде это отображается в циклы с небольшим телом и значительным числом итераций. При выборе расширения набора инструкций должны учитываться накладные расходы на декодирование инструкции и организацию циклов, длительность исполнения инструкции и возможность их переупорядочивания. Эти накладные расходы могут быть сравнимы с затратами на сами вычисления. В случае значительных накладных расходов для эффективной реализации алгоритма вычислительный блок необходимо разрабатывать не как расширение набора инструкций, а как акселератор вычисления библиотечных функций, поскольку в этом случае он не взаимодействует с конвейером процессора, и накладные расходы сводятся к минимуму.

Существует два основных подхода к разработке энергоэффективных схем с точки зрения выбора оптимального параллелизма вычислений и тактовой частоты:

- «равномерная работа», когда производительность и тактовая частота подбираются в соответствии с вычислительной нагрузкой;
- «рывок ко сну», когда устройство функционирует с максимальной эффективной частотой и по окончании работы засыпает с отключением питания от части блоков.

Критическим отличием малопотребляющих литографических процессов с малыми геометрическими нормами (45~нм и менее) от более старых является доминирование потерь мощности из-за токов утечки над активной мощностью, связанной с переключением логических со-

стояний схемы при работе в низкоэнергетических режимах с пониженной тактовой частотой. Токи утечки фактически пропорциональны площади подключенной к питанию схемы на кристалле. Подход «рывок ко сну» может являться более предпочтительным для таких схем по нескольким причинам:

- площадь блоков памяти, как правило, доминирует над площадью вычислительных схем, при этом размер памяти является характеристикой алгоритма и слабо зависит от скорости вычислений;
- на время простоя возможно полное отключение от питания блоков включая часть памяти для экономии энергии;
- не требуется динамическое управление тактовой частотой в низкоэнергетических режимах, что упрощает процесс разработки;
- устройство имеет запас вычислительных ресурсов, что увеличивает возможности модернизации и переиспользования без изменения архитектуры.

Базовыми алгоритмическими блоками для алгоритмов цифровой обработки сигналов являются сумматоры, умножители и запоминающие устройства. Для сумматоров и умножителей основным способом уменьшения энергопотребления является уменьшение количества используемых вентилей. Для реализации сумматоров используется троичная арифметика и древовидная структура вычислений [2]. Это позволяет избежать длинных цепочек переносов, что приводит к выравниванию путей и уменьшению задержки на критическом пути. Средства логического синтеза выполняют автоматические оптимизации графов вычислений из сумматоров и умножителей для достижения оптимальной площади схемы при заданной целевой частоте, которое достигается при выравнивании длины путей в схеме. Ручные оптимизации на уровне базовых операций как правило приводят к худшим результатам по энергопотреблению по сравнению с автоматическими оптимизациями в средствах логического синтеза.

На следующем уровне располагаются параллельные векторные операции над малыми векторами, векторное сложение, умножение, операции над кватернионами, линейная фильтрация. Энергоэффективная реализация этих операций не представляет сложностей с алгоритмической точки зрения, однако требует выбора и учета специфики архитектуры памяти, для обеспечения необходимого уровня параллелизма без дополнительных затрат энергии на одну операцию по сравнению с последовательной реализацией.

Общая модель энергопотребления тактируемых логических схем

Рассмотрим факторы, влияющие на энергоэффективность вычислительных блоков при построении малопотребляющих полупроводниковых схем, специализированных для группы алгоритмов. Будем учитывать и другие два дополнительных критерия – скорость разработки и удобство композиции вычислительных блоков для реализации сложных алгоритмов.

К физическим факторам относятся пороговое напряжение, токи утечки, паразитная емкость. Другими факторами являются архитектура, тип и размер памяти, точность промежуточных вычислений, методы логического синтеза. Проанализируем взаимосвязь этих факторов и их влияние на энергопотребление.

Поскольку рассматриваемые схемы работают непрерывно, то для характеризации их энергопотребления используется мощность. Мощность полупроводниковой схемы складывается из активной и статической рассеиваемой мощности. $P = P_a + P_s$.

Активная мощность обусловлена переключением состояния полупроводниковой схемы и связана с интенсивностью вычислений. Статическая мощность не зависит от вычислений, а только от количества аппаратуры, подключенной к питанию.

Сложные вычислительные блоки разрабатываются в виде тактируемых полупроводниковых схем. Схема состоит из регистров, хранящих состояние, и комбинационных логических схем, описывающих переход между состояниями. Переход между состояниями происходит по стробу тактирующего сигнала. Такая схема может быть представлена как конечный автомат.

$$F = (R, \sigma, \Sigma); \Sigma \subset R \subset \{0,1\}^n; \qquad \sigma : R \mapsto R,$$

где R - множество допустимых состояний, Σ - множество начальных состояний, - функция перехода. Функция перехода в полупроводниковой схеме реализуется как направленный ациклический граф элементарных библиотечных функций или вентилей. Вентили могут реализовывать такие функции как AND, OR, NOR, XOR, мультиплексор, селектор и т.д. На более низком уровне вентили состоят из транзисторов. Состояние системы хранится в регистрах. Такой уровень детализации описания логических схем называется уровнем регистровых передач (Register Transfer Level).

Мощность КМОП устройств

Для оценки мощности будем опираться на базовую модель, описанную во второй главе [3]. Для одного вентиля мощность определяется как

$$P = P_a + P_s$$

где P_a - активная мощность, зависящая от данных, P_s - статическая мощность, не зависящая от данных.

$$P_a = P_d + P_{sc}$$
$$P_s = P_{leak},$$

где P_d - динамическая мощность, P_{leak} - потери мощности от токов утечки, P_{sc} - потери мощности в результате короткого замыкания при переключении транзисторов. Динамическая мощность оценивается как

$$P_d \simeq \alpha C V^2 f, \quad (1)$$

где C - паразитная емкость, V - напряжение питания, f - тактовая частота, α - коэффициент частоты переключения, показывающий среднюю вероятность открытия вентилей на каждом такте. Для сигнала строба $\alpha = 1$, для других в среднем 0.1. Максимальная тактовая частота работы схемы при данном напряжении питания связана с задержкой на критическом пути, то есть цепочке вентилей, имеющей самую большую задержку переключения.

$$f \le 1 / d_{crit}, \qquad d_{crit} = \sum d_i$$

Задержка на одном вентиле также связана с напряжением по следующей формуле:

$$d(V) \propto \frac{C_L V}{(V - V_T)^2} \simeq \frac{C}{V}, \qquad V >> V_T, \quad (2)$$

где C_L - емкость нагрузки, которая состоит из перезаряжаемых входом емкостей затворов транзисторов и паразитных емкостей проводов.

$$C_T = C_{ox} W L$$

где C_T – емкость затвора транзистора, V_T – пороговое напряжение, C_{ox} – удельная емкость оксида затвора, W, L - ширина и длина затвора транзистора.

Из-за дефектов изготовления цифровые схемы работают неустойчиво при напряжениях, близких к V_T; существует минимальное напряжение $V_0 >> V_T$, обеспечивающее устойчивую работу схемы. При этом напряжении схема может работать на максимальной тактовой частоте $f_0 = 1 / d_{crit}(V_0)$. Схема может работать и при меньшей тактовой частоте, если скорость ее работы достаточна для решения задачи. На интервале частот $f \in (0, f_0]$ рост мощности от частоты линеен. При больших тактовых частотах

$$V \propto \frac{1}{d_{crit}} \propto f, \qquad f > f_0.$$

Таким образом, объединяя эти две оценки, получаем оценку

$$P_d(f) \propto \begin{cases} f, & f \le f_0 \\ f^3, & f > f_0 \end{cases}$$

Потери мощности в результате короткого замыкания при переключении транзисторов определяются по следующей формуле:

$$P_{sc} \simeq \alpha T_{sc} I_{sc} V, \qquad T_{sc} = \frac{T_r + T_f}{2},$$

где T_r, T_f - время роста и спада сигнала соответственно. Для логических схем P_{sc} мал и может не учитываться. Мы не будем его учитывать в дальнейших расчётах.

Статическая мощность обусловлена в основном токами утечки, протекающими через вентили в закрытом состоянии и определяется для одного вентиля при комнатной температуре как

$$P_{leak} \simeq V I_{sub}, \qquad I_{sub} \propto C_{ox} W / L. \tag{3}$$

Таким образом, при уменьшении геометрических норм токи утечки растут. Наилучшая энергоэффективность достигается при $V = V_0$. Пороговое напряжение V_T увеличивается с увеличением длины затвора L, что уменьшает ток утечки.

В библиотеке компонентов для одной технологии могут быть вентили с различной геометрией и различным пороговым напряжением V_T. Увеличение порогового напряжения приводит к уменьшению быстродействия.

Типичная архитектура акселератора

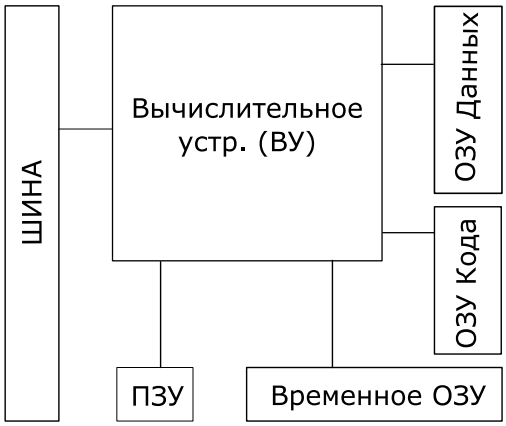


Рис. 1.

Типичная архитектура современных систем на кристалле включает множество специализированных компонент, объединенных одной или несколькими шинами или коммутатором. Каждый специализированный блок также обычно имеет типичную структуру, изображенную на рис. 1. Мощность такого блока складывается из статической мощности неотключаемых от питания компонент и мощности компонент с управляемым питанием. К неотключаемым компонентам относятся память кода и память состояния, остальные компоненты можно отключать на время простоя. К отключаемым компонентам относятся вычислительные устройства, память промежуточных данных и постоянное запоминающее устройство. Использование в постоянно работающих блоках энергонезависимой памяти на сегодняшний момент не оправдано из-за высоких накладных расходов на чтение и запись.

Оценка рассеиваемой мощности акселератора при фиксированном размере

Предположим, что фиксированный алгоритм выполняется на вычислительном блоке, в котором могут быть задействованы параллельно p процессоров. Требуется найти p, при котором потребляемая мощность блока минимальна.

Обрабатываемый алгоритм характеризуется следующими постоянными параметрами: B - вычислительная сложность алгоритма, выраженная в количестве элементарных операций в се-

кунду. Элементарная операция запускается на исполнение каждый такт. Обычно операции выполняются в конвейере, то есть время выполнения операции составляет несколько тактов. K - нераспараллеливаемая доля алгоритма, равная отношению длины критического пути к длине всего алгоритма.

Вычислительный блок характеризуется следующими постоянными параметрами: N_0 - количество вентилей в неотключаемой части, N_1 - количество вентилей в отключаемой части без вычислительных компонент, $\hat{N}_2$ - количество вентилей в последовательной реализации вычислительных компонент блока, f - тактовая частота.

Количество вентилей в вычислительных компонентах блока с учетом параллелизма определяется как $N_2 = p\hat{N}_2$. В этих обозначениях, используя формулы (1, 3), мощность вычислительного блока на участке линейного роста от частоты можно оценить как

$$P_s = P_{leak} = c_0(N_0 + S(N_1 + N_2))$$
$$P_a = P_d = c_1 f S(N_1 + N_2)$$
$$P = P_a + P_s = c_0(N_0 + S(N_1 + N_2)) + c_1 f S(N_1 + N_2),$$

где c_0, c_1 - постоянные коэффициенты, а $S = S(p) \le 1$ - скважность или отношение времени вычислений к общему времени, включающему простой.

При $p = 1$ скважность равна $S = B / f$ по определению. При $p > 1$ функцию $S(p)$ заменяют на функцию ускорения от параллелизма $l(p)$ в соответствии с уравнением

$$S(p) = \frac{B}{l(p) f}.$$

Функцию $l(p)$ можно описать с помощью закона Амдала [4], который используется в области параллельного программирования для оценки ускорения при фиксированном размере задачи

$$l(p) = \frac{p}{1 + K(p-1)}.$$

Таким образом,

$$P = c_0 N_0 + B(c_0 / f + c_1)\frac{p\hat{N}_2 + N_1}{l(p)} \quad (4)$$

При этом мощность достигает минимума на максимальной частоте линейного масштабирования f_0:

$$\arg\min_{f, f \le f_0} P = f_0.$$

Вопрос оптимального выбора параллелизма для минимизации энергопотребления при моделировании ускорения с помощью закона Амдала рассматривался Ву и Ли [5] для многоядерных суперскалярных процессоров. В отличие от рассматриваемой нами задачи, процессоры работают на высокой частоте и при высоком напряжении питания, что позволяет не учитывать энергопотребление памяти.

Лемма

Мощность достигает минимума по p при значении

$$p_0 = \arg\min_p P = \max\left(1, \sqrt{\frac{N_1(1-K)}{\hat{N}_2 K}}\right).$$

При этом минимальная мощность имеет значение

$$P_{\min} = c_0 N_0 + B(c_0 / f_0 + c_1)\left((1-K)\hat{N}_2 + KN_1 + 2\sqrt{KN_1\hat{N}_2(1-K)}\right).$$

Доказательство

Минимизируемую функцию мощности можно представить в виде

$$P(p) = \alpha p + \beta + \gamma p^{-1}$$

с подходящими коэффициентами α, β, γ. Дифференцированием находим условие минимума $\alpha - \gamma p^{-2} = 0$. Отсюда тождественными преобразованиями получаем заключение леммы. ■

Важным показателем качества распараллеливания является максимальная экономия энергии

$$\rho = \frac{P(1)}{P_{\min}},$$

где $P(1)$ есть мощность блока при отсутствии распараллеливания.

Следствие

При полном распараллеливании (при $K = 0$)

$$\rho = 1 + \frac{B(c_0 / f_0 + c_1)N_1}{P_{\min}}$$

$$P_{\min} = c_0 N_0 + B(c_0 / f_0 + c_1)\hat{N}_2. \square$$

В инженерной практике работы по распараллеливанию алгоритма имеет смысл проводить, только если оценка экономии энергии $\rho \geq 1.2$, поскольку алгоритмы логического синтеза схемы могут вносить непредсказуемый вклад такого же порядка в энергопотребление.

Таким образом, в рамках данной архитектуры для хорошо распараллеливаемых алгоритмов с доминированием памяти промежуточных результатов повышение параллелизма ведет к экономии энергии. Экономия достигается за счет уменьшения статической мощности памяти промежуточных данных с помощью отключения от питания.

Асимптотическая скорость роста мощности при росте размера задачи

Кроме задачи оценки оптимального параллелизма, рассмотрим задачу оценки роста мощности при увеличении размера задачи. Хотя обычно размер задачи в алгоритмах для низкоэнергетических устройств известен заранее, но в некоторых случаях его можно варьировать. Например, размер фильтров, количество оцениваемых параметров, разрешение изображений и т.п.

Функция мощности зависит от параллелизма p и размера задачи n:

$$P(p,n) = c_0 N_0(n) + B(n)(c_0 / f + c_1)\frac{p\hat{N}_2 + N_1(n)}{l(p,n)},$$

где величины $N_0(n)$, $N_1(n)$ зависят от размера памяти, который обычно растет с ростом размера задачи. Величина $\hat{N}_2$ не зависит от n, поскольку алгоритм вычислений не меняется.

Для оценки $l(p,n)$ используем закон Густафсона-Барсиса [6], который используется в области параллельного программирования для оценок при росте размера задачи:

$$l(p,n) = (1-\alpha)p + \alpha.$$

Закон Густафсона-Барсиса, предполагает, что непараллелизуемая часть задачи не растет с ростом размерности, что часто встречается при независимости параллельных подзадач.т друга.

Лемма

Пусть размер памяти $N_0(n)$, $N_1(n)$ и сложность алгоритма $B(n)$ растут линейно по размеру задачи. Тогда при отсутствии распараллеливания мощность $P(1,n) = O(n^2)$, а минимальная мощность $P_{\min}(n) = O(n)$ при $n \to \infty$.

Доказательство

Подставим линейные функции и закон Густафсона-Барсиса:

$$P(p,n) = c_0 N_0 n + Bn(c_0 / f + c_1)\frac{p\hat{N}_2 + N_1 n}{(1-\alpha)p + \alpha}.$$

При отсутствии распараллеливавния $p = 1$. Очевидно, что в этом случае величина $P(1,n)$ имеет порядок n^2 при $n \to \infty$. Если коэффициент распараллеливания $p = p(n)$ пропорционален n, то $P(p(n),n) = O(n)$.

Влияние перспективных технологий на энергоэффективность

Производство полупроводниковых схем является быстро развивающейся областью. Качественные оценки энергоэффективности и выбор архитектуры привязаны к поколению технологических процессов и к специфике технологических библиотек компонентов. При планировании перспективных разработок следует учитывать не только текущее состояние, но и перспективные технологии, планируемые к внедрению в производство в течение ближайших нескольких лет. Такими технологиями для малопотребляющих технологических процессов на сегодня являются:

- Схемы с пороговым напряжением питания (Near-threshold voltage circuits);
- Память на основе фазового перехода (Phase Change Memory);
- Магниторезистивная память (MRAM).

Схемы с пороговым напряжением питания

Для устойчивой работы цифровых схем напряжение питания V должно в несколько раз превышать пороговое напряжение транзисторов V_T. Уменьшение напряжения питания ведет к уменьшению как активной, так и статической рассеиваемой мощности схемы. При этом по формуле (2) максимальная тактовая частота уменьшается пропорционально

$$f_0 \propto \frac{(V - V_T)^2}{V}.$$

Однако следует помнить, что схема должна выполнять некоторую фиксированную вычислительную нагрузку. Если схема не может выполнять эту нагрузку, то ей требуется распараллеливание, обратно пропорциональное уменьшению тактовой частоты, что ведет к потенциально бесконечному увеличению площади и мощности. Оценим энергию на выполнение одной операции при частоте f_0.

$$E_{f_0} \propto V^2 \left(\alpha C + \beta \frac{I_{sub}}{(V - V_T)^2} \right) \lim_{V \to V_T} E_{f_0} = \inf$$

Уменьшение напряжения питания логических схем ведет к увеличению вероятности ошибки, для предотвращения ошибок требуется использовать сквозное линейное кодирование с исправлением ошибок и/или многократное вычисление результатов с голосованием, что уменьшает выигрыш от понижения напряжения. Кроме того, такое изменение подхода к проектированию не поддерживается инструментальными средствами разработки, что ведет к увеличению трудоемкости разработки специализированных схем для набора алгоритмов.

При использовании подхода ``рывок ко сну'' энергопотребление в основном обусловлено токами утечки памяти состояния. Мы можем заключить, что по этим причинам для вычислительных блоков использование схем с пороговым напряжением питания в ближайшем будущем не будет оправдано в малопотребляющих специализированных схемах. Это соответствует текущему позиционированию технологии для микропроцессоров и микроконтроллеров общего применения и медицинского, в том числе имплантируемого оборудования. Поскольку микропроцессоры общего применения выпускаются большими партиями, расходы на разработку

амортизируются в структуре себестоимости. Имплантируемое оборудование является функционально простым и дорогостоящим, что тоже позволяет использовать менее технологичные методы разработки.

Другим способом использования технологии является разработка памяти с пониженным напряжением питания в режиме хранения данных. Использование такой памяти не представляет сложности с точки зрения инструментальных средств разработки, поскольку она является крупным универсальным библиотечным блоком и подключается через стандартный интерфейс. Уменьшение напряжения питания памяти во время простоя существенно уменьшит мощность схем при подходе ``рывок ко сну".

Для отражения влияния памяти с пониженным напряжением питания в режиме простоя в формуле (4) следует заменить член $c_0 N_0$ на сумму $c_0 N_0 + c_0^* N_0^*$ где второе слагаемое отвечает за статическую мощность памяти. Формула принимает следующий вид,

$$P = c_0 N_0 + c_0^* N_0^* + B(c_0 / f + c_1)\frac{(r(p)\hat{N}_2 + N_1)}{l(p)},$$

что не меняет ее форму.

Память на основе фазового перехода и магниторезистивная память

Другими перспективными видами памяти являются память на основе фазового перехода и магниторезистивная память. Эти виды памяти изменяют сопротивление ячеек при переключении состояния. Так же, как и флэш-память, эти типы памяти являются перезаписываемыми, то есть не потребляют энергии при хранении данных. Они могут быть сформированы на кристале в рамках того же технологического процесса, что и логические схемы. В отличие от NAND флэш-памяти они имеют произвольный доступ и большое количество циклов перезаписи, сравнимое с оперативной памятью. Энергия чтения сравнима с оперативной памятью, однако энергия записи в несколько раз превосходит энергию чтения.

Эти виды памяти хорошо подходят для хранения таблиц и кода программы. Для отражения влияния новых типов памяти изменения в формуле (4) не требуются. Достаточно учесть разность стоимости чтения и записи в множителе B, отвечающем за сложность алгоритма.

Заключение

В работе рассматривается задача оптимизация энергопотребления малопотребляющих устройств с реконфигурируемой вычислительной архитектурой на основе графов потоков данных, часто называемых антимашинами Хартенштейна с учетом активной мощности и токов утечки. Предлагается модель энергопотребления при варьировании скважности вычислений. В отличие от обычных моделей энергопотребления на основе подсчета элементарных операций в этой модели естественным образом возникает задача нахождения оптимального параллелизма при проектировании устройства. Эта задача решается в работе. Параллелизм устройства существенно влияет на энергопотребление устройства. Результаты могут использоваться для практического проектирования малопотребляющих полупроводниковых схем для цифровой обработки сигналов.

ЛИТЕРАТУРА

1. Hartenstein R. W. et al. A novel ASIC design approach based on a new machine paradigm //IEEE Journal of Solid-State Circuits. – 1991. – Т. 26. – №. 7. – С. 975-989.
2. Earle J. G. Latched carry save adder circuit for multipliers : пат. 3340388 США. – 1967.
3. Panda P. R. et al. Power-efficient system design. – Springer Science & Business Media, 2010.
4. Amdahl G. M. Validity of the single processor approach to achieving large scale computing capabilities //Proceedings of the April 18-20, 1967, spring joint computer conference. – ACM, 1967. – С. 483-485.
5. Woo D. H., Lee H. H. S. Extending Amdahl //Computer. – 2008. – №. 12. – С. 24-31.
6. Gustafson J. L. Reevaluating Amdahl's law //Communications of the ACM. – 1988. – Т. 31. – №. 5. – С. 532-533.

S.I.Salishev (Saint-Petersburg University, Saint-Petersburg)

Implementation and estimation of power consumption of low-power semiconductor circuits for digital signal processing based on Hartenstein anti-machines

An approach to the development of low-power semiconductor logic circuits oriented at performing efficient parallel computations typical for digital signal processing based on non-processor architectures consisting of large multifunctional computing units connected by a reconfigurable graph, also called Hartenstein's anti-machines, is considered.A model of energy consumption of such devices is proposed. The problem of choosing the optimal parallelism for the device is solved.

АДАПТАЦИЯ ФУНКЦИИ КОРРЕЛЯЦИИ КАНАЛОВ В МИКРОФОННЫХ РЕШЕТКАХ ДЛЯ УЛУЧШЕНИЯ КАЧЕСТВА РАСПОЗНАВАНИЯ РЕЧИ ПРИ ПОМОЩИ ПОСТФИЛЬТРА ШУМОПОДАВЛЕНИЯ

Е. И. Викулов
(*Ланит-Терком, Россия, 198504, С.-Петербург, Петергоф, Университетский пр., 28, +7 (812) 428-74-09, contact@lanit-tercom.com*)

С.И. Салищев
(*СПбГУ, Россия, 199034, С.-Петербург, Университетская набережная 7–9, +7 (812) 328–20–00, spbu@spbu.ru*)

Целью работы является разработка алгоритма адаптации функции корреляции каналов в микрофонных решетках, вызванной дифракцией и реверберацией звуковых волн шумового сигнала на окружении. Более точная оценка корреляционной функции приводит к улучшению работы постфильтра и повышению качества распознавания речи в выходном сигнале постфильтра.

В работе рассматривается подход к шумоподавлению в микрофонных решетках на основе постфильтра. Одним из наиболее сложных сценариев приема сигнала с помощью микрофонной решетки является случай приема направленного полезного сигнала на фоне изотропного, т. е. присутствующего во всех пространственных каналах, шума. Такой шум получается в результате дифракции и реверберации звуковых волн на окружении и характерен для звуковых волн в газах и жидкостях из-за низкой скорости звука в этих средах. Поэтому данный сценарий приема в основном изучается в акустических приложениях. Микрофонная решетка позволяет применять несколько различных диаграмм направленности одновременно, что можно использовать для оценки направленного сигнала, подавляя пространственные каналы, не содержащие полезный сигнал и оценки изотропного шума, наоборот подавляя пространственный канал, содержащий полезный сигнал. При наличии оценки отношения сигнал/шум возможно применение различных линейных или нелинейных фильтров для подавления изотропного шума, называемых постфильтрами.Подход основан на применении фильтра Винера с использованием пространственной информации для оценки отношения сигнал шум, и был предложен Зелинским для некоррелированного шума[1]и обобщен Маккованом[2] а случай коррелированного шума. В работе Лефкимиатиса и Марагоса[3] рассмотрен случай логарифмически оптимальных фильтров более предпочтительных для систем обработки и воспроизведения звука. В работе Салищева, Барабанова и Клочкова[4] рассматриваются гребенчатые постфильтры на основе классификации типа сигнала, улучшающие работу систем автоматического распознавания речи.

Дифракция звуковой волны на обрамлении микрофонов оказывает большое влияние на корреляционную функцию каналов, от которой зависит оценка отношения сигнал/шум. Например, встраивание микрофонов в сферическую оболочку приводит к эффекту аналогичному увеличению расстояния между микрофонами в 3 раза по сравнению со свободно стоящими микрофонами для голосовой полосы звуковых частот. Также на корреляционную функцию влияет, находятся ли микрофоны в помещении или вне помещения, и звукопоглощающие свойства материалов окружения.Для помещений наблюдаемая корреляционная функция соответствует сферически изотропному шуму, для открытых пространств с отражением от поверхности земли – цилиндрически изотропному шуму[5].

Поскольку все эти факторы крайне сложно или невозможно учесть заранее, система шумоподавления на основе постфильтрадолжна включать механизм адаптации корреляционной функции каналов. Основной идеей алгоритма является выделение в сигнале сегментов, не имеющих большой корреляции, т.е. не содержащих направленный сигнал и не содержащих квазигармонического сигнала, например, речи. Эти сегменты используются для оценки корреляции каналов, которая затем усредняется с помощью фильтра.

Система с адаптивным алгоритмом подстройки корреляционной функции демонстрирует более высокое качество распознавания по сравнению с той же системой в предположении о некоррелированности каналов и при несовпадении типа окружения и априорной оценки изотро-

пии шума, например, использовании в помещении при априорной оценке вне помещения.С использованием системы распознавания речи Sphinx 4[6] и гребенчатого постфильтра[4], улучшение точности распознавания составило до 40% при использовании адаптивной оценки корреляции по сравнению с функцией корреляции сферически изотропного шума.

Основными результатами работы являются алгоритм адаптации корреляционной функции и оценки качества распознавания речи с/без использования алгоритма, демонстрирующие улучшения качества распознавания в случае несовпадения истинной функции корреляции с ее априорной оценкой.

ЛИТЕРАТУРА

1. Zelinski R. A microphone array with adaptive post-filtering for noise reduction in reverberant rooms //Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on. – IEEE, 1988. – С. 2578-2581.
2. McCowan I. A., Bourlard H. Microphone array post-filter based on noise field coherence //IEEE Transactions on Speech and Audio Processing. – 2003. – Т. 11. – №. 6. – С. 709-716.
3. Lefkimmiatis S., Maragos P. A generalized estimation approach for linear and nonlinear microphone array post-filters //Speech communication. – 2007. – Т. 49. – №. 7-8. – С. 657-666.
4. Salishev S., Klotchkov I., Barabanov A. Microphone Array Post-filter in Frequency Domain for Speech Recognition Using Short-Time Log-Spectral Amplitude Estimator and Spectral Harmonic/Noise Classifier //International Conference on Speech and Computer. – Springer, Cham, 2017. – С. 525-534.
5. Cox H. Spatial correlation in arbitrary noise fields with application to ambient sea noise //The Journal of the Acoustical Society of America. – 1973. – Т. 54. – №. 5. – С. 1289-1301.
6. Lamere P. et al. The CMU SPHINX-4 speech recognition system //IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2003), Hong Kong. – 2003. – Т. 1. – С. 2-5.

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ ПРИМЕНЕНИЯ ПЕРСПЕКТИВНЫХ ОТЕЧЕСТВЕННЫХ ВЕКТОРНЫХ ПРОЦЕССОРОВ ПРИ ОБРАБОТКЕ СИГНАЛОВ И МНОГОМЕРНОЙ ВИЗУАЛИЗАЦИИ

Д. В. Богаевский[1], **С. Н. Ежов**[2], **Д. И. Каплун**[3], **А. Д. Кошкин**[4]
(*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина), С.-Петербург, ул. Профессора Попова, 5, dikaplun@etu.ru*)

С. И. Аряшев[5]
(*Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук, Москва, Нахимовский просп., 36, к.1, aserg@cs.niisi.ras.ru*)

Аннотация

В работе выделен минимальный набор команд векторного сопроцессора, позволяющего эффективно выполнять расчеты с вещественной арифметикой в задачах обработки сигналов и изображений. Произведены: выбор методов профилирования; подготовка данных и модификация эмулятора; профилирование по различным параметрам; процедуры анализа и проверка результатов, позволяющие оценить параметры алгоритмов для эмулируемого процессора. Разработана методика использования в системе команд процессора целочисленной арифметики малой разрядности (8, 16, 32 разряда) для эффективного выполнения традиционных пользовательских задач, таких как кодирование/декодирование видео и аудио данных, обработка статичных изображений. Результаты работы могут быть использованы при построении высокопроизводительных систем обработки изображений, встраиваемых систем анализа многомерных сигналов, при разработке суперкомпьютеров, предназначенных для имитационного моделирования сложных естественных и технических процессов.

Ключевые слова: векторизация, процессор, визуализация, SIMD, QEMU.

Введение

Геополитические реалии современного мира формируют острую потребность в разработке отечественных микропроцессоров, способных обеспечивать высокую производительность научно-технических расчётов, а также имитационного моделирования, обработки сигналов и изображений даже в экстремально жёстких условиях эксплуатации. Для удовлетворения этой потребности в работе будет предложена архитектура семейства высокопроизводительных микропроцессоров, в которых в качестве расширения базовой архитектуры включены специализированные сопроцессоры, ориентированные (согласно принципу “встречной оптимизации”) на ускоренное исполнение заданного набора математических функций, наиболее часто употребляемых при решении задач из обозначенных выше областей применения.

Все современные процессоры в том или иной форме поддерживают векторную арифметику на аппаратном уровне. Правильное использование векторных регистров позволяет получить существенный прирост производительности благодаря выполнению нескольких однотипных операций за одну инструкцию. Основным методом решения ряда задач обработки изображений, многомерных сигналов, формирования объектов трехмерной графики является использование векторных процессоров и многопоточных вычислений. Актуальность данной работы заключается в исследовании возможностей использования современного отечественного процессора для обработки изображений, графических объектов и многомерного анализа сигналов.

Основным принципом повышения производительности вычислений является их параллельное исполнение, используемое на всех уровнях организации вычислительного устройства, архитектуры команд и непосредственно микроархитектуры процессора. Следует отметить, что простое увеличение используемых элементарных конструкций, таких как лингвистические единицы языка программирования, количества элементов (АЛУ, ядер) практически никогда не

[1] Аспирант.
[2] Кандидат технических наук, доцент.
[3] Кандидат технических наук, доцент.
[4] Студент.
[5] Кандидат технических наук, начальник отделения.

приводит к пропорциональному росту полной вычислительной производительности. Предсказание поведения приложений на будущих, ещё не спроектированных системах, представляется крайне важным элементом работы, так как позволяет заранее обнаружить слабые места проекта и перераспределить средства для получения оптимальной системы. Целью работы является полноплатформенная симуляция вычислительных векторных систем с их последующим прототипированием на FPGA, решающая задачу прогнозирования производительности и иных характеристик аппаратных решений до запуска серийного производства.

Для построения и исследования моделей был использован программный симулятор QEMU (www.qemu.org), предоставляющий ряд инструментов, необходимых для проводимого исследования: исполнение графических приложений на всех процессорах системы, распределение нагрузки как на отдельные элементы архитектуры команд, так и все узлы вычислителя, изучение влияния немодифицированной операционной системы. Включение всех деталей работы векторного микропроцессора, устройств оперативной памяти, сетевых маршрутизаторов приводит к существенному замедлению и невозможности проведения значительного объёма исследований в разумное время. Поэтому нами был использован гибридный подход, в котором моделирование используется для получения значений ключевых параметров, характеризующих работу приложения внутри симулируемой системы. Полученные статистические данные в дальнейшем будут применены в формульных описаниях и оценках производительности векторных многопроцессорных вычислителей при решении различных классов задач.

Современные подходы к повышению производительности процессоров при обработке и визуализации сигналов

Над возможностью векторизации для повышения производительности процессора при выполнении определённых классов задач работало множество исследователей. Callahan в работе [1] исследовал начальный набор микротестов для оценки возможностей векторизации. Тесты были написаны на языке Фортран, и тематическое исследование выполнялось на ранних семействах SIMD в рамках архитектуры компьютеров общего назначения. Maleki и др. [2] расширили оригинальный набор тестов с поддержкой языка С, и добавили еще несколько микротестов. В основу данного проекта положен подход, предложенный Maleki, с дополнением в виде ряда ослабленных эвристик, предназначенных для адекватной векторизации класса задач графической визуализации путем добавления соответствующих критериев, в частности, относящихся к рекуррентному анализу сигналов и построению динамических многомерных карт отображений.

Существующие на сегодняшний день программные пакеты и библиотеки для визуализации динамики систем и процессов [3,4,5] не предлагают эффективных решений, учитывающих специфику задач, и обуславливают возникновение ряда сложностей при построении программно-аппаратных комплексов АСНИ на основе отечественных векторных вычислительных устройств.

Первая сложность заключается в многомерности процесса пространственного мониторинга, обуславливающего существенные трудности при формировании визуальных образов получаемых результатов, особенно там, где оператор не обладает априорной информацией о характере нестационарных взаимодействий между различными факторами. Существующие системы пространственного мониторинга представляют собой «черный ящик» с фиксированными взаимосвязями между различными факторами, в которых без применения средств визуального представления и контроля, получаемых в процессе анализа результатов, даже опытному оператору трудно представить, что происходит в области мониторинга. Для представления решения поставленной задачи многомерной мониторинга необходимо использование комбинированных подходов графического представления, что позволит обеспечить качественное понимание результирующей картины. Следует заметить, что в современных средствах используется только двумерная индикация сигнальной информации [6]. Это приводит к тому, что на экране отображается значительное множество отдельных двумерных объектов, соответствующих проекциям многомерного сигнала на плоскость. Подобное представление затрудняет формирование многомерного образа оператором, увеличивает время принятия решений, что является очень суще-

ственным в реальных тактических ситуациях. Трехмерная визуализация снижает количество необходимых для контроля оператором модулей, при этом повышается информативность остальных модулей благодаря возможности размещения большего количества информации. Предлагаемое решение задач визуализации в составе средств систем пространственного мониторинга позволяет согласовать алгоритмы визуального представления результатов мониторинга с законами физических процессов. Вторая сложность заключается в отсутствии кроссплатформенных решений, которые могли бы работать под управлением отечественных операционных систем для высокопроизводительных векторных вычислительных комплексов, построенных на отечественной элементной базе (например, микропроцессоры с архитектурой КОМДИВ64, Эльбрус 401-РС и др.).

Henretty и др. [7] удалось улучшить векторизацию с использованием шаблонов решений, но не удалось распространить эти подходы на другие типы ядер. Nuzman [8] несколько улучшил производительность векторизации внешнего цикла на архитектурах CBE и PowerPC.

Holewinski [9], Evans [10] и Barrik [11] использовали информацию о трассировке для определения потенциала возможностей векторизации и автоматизации выбора векторных инструкций. Hohenauer [12] предложил структуру, позволяющую компиляторам с возможностью переназначения выдавать более подходящие инструкции для SIMD. Mc Farlin и др. [13] представили автоматическую векторизацию ядра, выполняющего БПФ. Обе эти работы были ориентированы на модификации компилятора для выполнения оптимизаций и не могут быть напрямую распространены на новые архитектуры или компиляторы.

При проектировании процессоров особое внимание уделяется способам представления данных, особенно если операции выполняются над целыми числами. Одним из таких способов является система остаточных классов. В настоящее время система остаточных классов привлекает большое внимание исследователей во всем мире. В монографиях [14] и [15] приведены основы теории системы остаточных классов, методы ее применения на практике, а также произведен детальный обзор проблемных направлений в данной научной области. Среди задач, стоящих перед исследователями в настоящее время, можно выделить следующие: разработка устройств на основе системы остаточных классов для практического применения и поиск наилучших наборов модулей системы остаточных классов для решения конкретных задач.

Главной особенностью системы остаточных классов является очень быстрое выполнение операций сложения, вычитания и умножения по сравнению с традиционной двоичной системой счисления. При этом выполнение операций сравнения чисел, деления, восстановления позиционной формы и некоторых других в данной системе достаточно проблематично. Поэтому в настоящее время активно ведутся исследования по практическому применению системы остаточных классов именно в тех приложениях, в которых основная доля вычислений приходится на операции сложения, вычитания и умножения. Докторская диссертация [16] посвящена применению системы остаточных классов в цифровых фильтрах, выполняющих преобразование Фурье. Ведущими зарубежными специалистами в области цифровой обработки сигналов с применением арифметики системы остаточных классов являются G.-C. Cardarilli, U. Meyer-Baese, F. Taylor, G. Jullien и K. Navi. В работах этих авторов убедительно доказывается преимущество системы остаточных классов перед двоичной системой счисления по скорости выполнения цифровой фильтрации. Несмотря на это, количество работ по цифровой обработке изображений с использованием системы остаточных классов весьма мало, по сравнению с количеством работ по цифровой обработке сигналов на базе системы остаточных классов в целом. Это связано со спецификой алгоритмов обработки изображений, требующих учета взаимного расположения обрабатываемых пикселей. В настоящее время разработаны методы увеличения резкости и выделения контуров изображения с использованием системы остаточных классов [17], методы сглаживания изображений на основе системы остаточных классов [18, 19]. Кроме того, разработана архитектура сопроцессора цифровой обработки изображений, выполняющая дискретное вейвлет-преобразование и использующая арифметику системы остаточных классов [20]. Авторы этой работы показали увеличение производительности на 7% и снижение аппаратных затрат на 5% по сравнению с аналогичным устройством, использующим традиционную двоичную систему счисления. Исходя из общего малого количества опубликованных

работ по цифровой обработке изображений с использованием системы остаточных классов, можно сделать вывод о новизне и перспективности данного научного направления.

Появление современных средств моделирования и проектирования цифровых устройств в виде связки FPGA-ASIC не только подтвердило давно известные достоинства системы остаточных классов в скорости выполнения арифметических операций и высокой способности к обеспечению отказоустойчивости, но и обнаружила ее новое качество - возможность снижения энергопотребления цифровых устройств. Снижение энергопотребления в цифровых устройствах на базе модулярной арифметики объясняется существенным уменьшением межразрядных переносов, что позволяет сократить затраты электроэнергии на поддержание уровней сигнала при выполнении арифметических операций. Однако в последних работах [21, 22] приводятся результаты моделирования лишь для отдельно взятых наборов модулей системы остаточных классов. При этом вопрос о важности оптимального выбора количества модулей, а также их вида, для применения в конкретном практическом приложении, в настоящее время стоит весьма остро [23]. По этой причине, одной из частных задач настоящего исследования является решение проблемы выбора оптимального набора модулей системы остаточных классов в системах цифровой обработки изображений.

Формирование требований к архитектуре и производительности отечественного процессора

Из вышесказанного следует, что целью работы было исследование и имитационное моделирование комплекса задач обработки сигналов, изображений и графических объектов для формирования перечня команд, значимых с точки зрения производительности и пригодных к реализации на предложенной архитектуре векторного сопроцессора. В соответствие с целью исследования были решены следующие задачи:

1. Формирование перечня команд, выполнение которых в значительной степени определяет производительность разрабатываемой системы на кристалле на исследуемых задачах.
2. Формирование паттернов программных решений для эффективного выполнения исследуемых задач на перспективной архитектуре векторного сопроцессора с поддерживаемыми командами из референсного MIPS SIMD Architecture (MSA) расширения.
3. Разработка тестовых программ и моделей основных задач и функций обработки изображений и видео для проверки функционирования модифицированной структуры векторного сопроцессора.
4. Сравнительная оценка производительности систем, реализованных в арифметике остаточных классов, с классическими подходами к обработке сигналов и изображений с использованием векторного сопроцессора.

Векторный сопроцессор

Предлагаемая архитектура векторного сопроцессора [24] используется в качестве расширения универсальных микропроцессоров для ускорения вычислений над векторами комплексных и вещественных чисел одинарной и двойной точности.

Структурная схема векторного сопроцессора приведена на рис. 1 [24]. Вычислительное ядро векторного сопроцессора позволяет выполнять вещественные операции одинарной и двойной точности над векторами шириной 128 разрядов. Пиковая производительность достигается при использовании команд комплексного умножения с прибавлением и вычитанием третьего операнда («бабочка Фурье») и составляет за такт 10 вещественных операций двойной точности или 20 вещественных операций одинарной точности. Отказ от использования локальной специализированной памяти позволяет использовать вычислительные возможности сопроцессора на пользовательских приложениях, скомпилированных стандартными средствами с подключением математических библиотек.

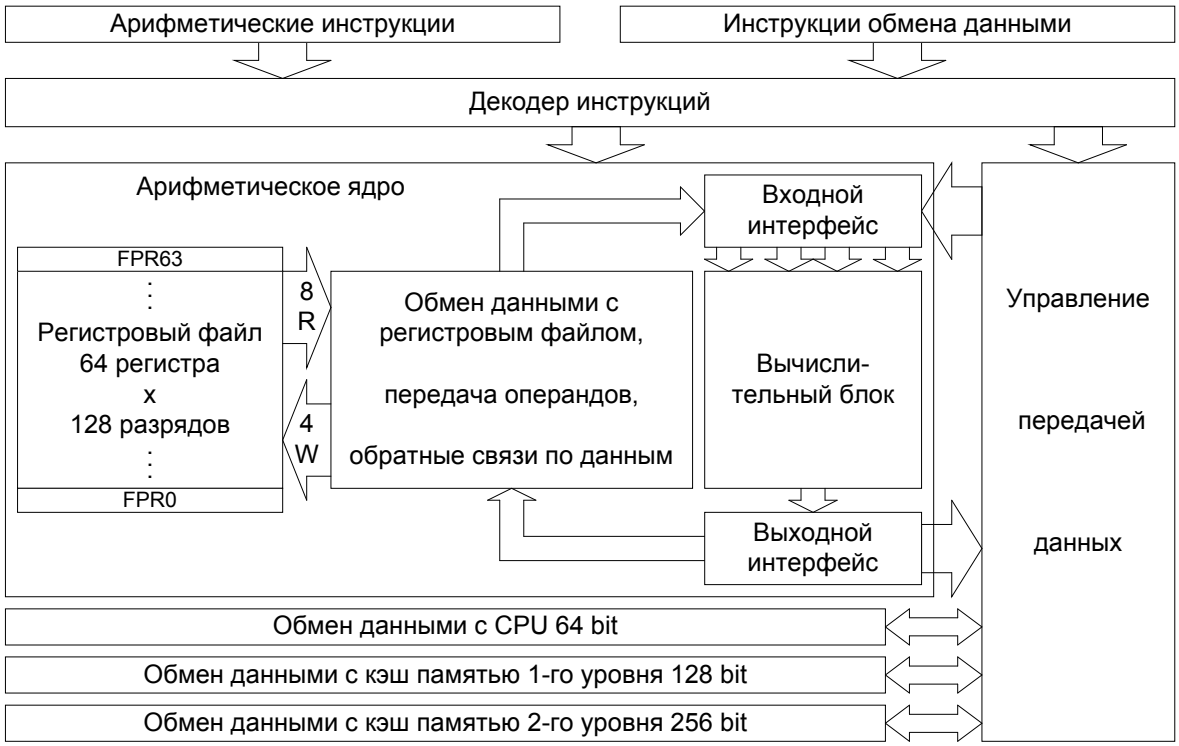


Рис. 1. Структурная схема векторного сопроцессора процессора

В составе исполнительных арифметических блоков, векторный сопроцессор содержит четыре модуля FMA (multiply-add fused), два модуля вычисления обратного квадратного корня и два модуля вычисления обратной величины. Необходимо отметить, что все из вышеперечисленных модулей выполняют операции над одним набором 64-разрядных входных операндов или над двумя наборами 32-разрядных входных операндов. Также присутствуют модули конвертации форматов арифметических данных и модули сравнения.

Система команд векторного сопроцессора позволяет эффективно выполнять инженерные расчеты с вещественной арифметикой. Так, при реализации функций BLAS, максимальное значение эффективности приближается к 64% для случая, когда исходные данные полностью помещаются в кэш память второго уровня, число арифметических команд достаточно велико, чтобы скрыть затраты на передачу данных, а комплексный тип данных позволяет использовать наиболее производительные команды сопроцессора.

Однако, для эффективного выполнения традиционных пользовательских задач, таких как кодирование/декодирование видео и аудио данных, обработка статичных изображений сопроцессору необходима поддержка целочисленной арифметики. Причем, необходима арифметика малой разрядности 8, 16 разрядов, реже 32 разряда.

Обзор основных составляющих задач обработки изображений и сигналов

В первую очередь были проведены исследования по оценке эффективности использования вычислительного ядра с целочисленным SIMD набором команд (MIPS MSA-совместимым) на следующем наборе задач обработки изображений:

1. ***Задачи фильтрации изображения методом свертки с окном.***
2. ***Задачи сжатия статичных/видео изображений, включающие в себя:***

• JPEG кодирование/декодирование (DCT-based encoding- ijg.org, wavelet encoding – jpeg2000).

• H264 кодирование/декодирование.

– основных функций обработки изображений:

• Фильтрация изображения методом свертки с окном.

• Преобразование цветовых пространств (RGB-YUV).

- Предварительная и постобработка FDCT и IDCT.
- Квантование/Деквантование, восстановление.
- Оценка и компенсация движения.
- Кодирование/декодирование VLC.

3. Задачи обработки 3D графики:

- Вращение, перемещение, масштабирование координат 3D объектов.
- Растеризация. Перевод изображения, описанного векторным форматом в пиксели.
- Raycasting. Метод рендеринга, в котором расчёт цвета каждой точки картинки построением луча из точки зрения наблюдателя через воображаемое отверстие в экране на месте этого пикселя «в сцену» до пересечения с первой поверхностью. Цвет пикселя будет таким же, как цвет этой поверхности (иногда с учётом освещения и т. д.).
- Raytracing. Метод рендеринга, в котором конечный цвет пикселя рассчитывается при помощи преломляющихся и отражающихся лучей, выпущенных из глаз наблюдателя сцены.
- Текстурирование. Текстурирование подразумевает проецирование растровых или процедурных текстур на поверхности трёхмерного объекта в соответствии с картой UV-координат, где каждой вершине объекта ставится в соответствие определённая координата на двухмерном пространстве текстуры.

Исследование возможностей векторизации в кодеке стандарта H.264 с применением контекстно-адаптивных кодов переменной длины (CAVLC)

В кодере H.264 метод CAVLC [25] имеет дело с просканированными по зигзагу блоками 4×4, 2×4 или 2×2 квантованных коэффициентов. Таким образом, на вход CAVLC подаётся ве ктор из 16, 8 или 4 элементов. На выходе CAVLC получается битовая строка, соответствующая кодируемому вектору. Обычно после предсказания, преобразования и квантования блоки коэффициентов становятся весьма разреженными, т.е. содержат много нулевых коэффициентов. Для компактного представления серий нулевых коэффициентов используется схема «серия-значение», которая в случае CAVLC подразумевает отдельное кодирование ненулевых коэффициентов и отдельное кодирование серий нулей перед ними.

Алгоритм при использовании CAVLC состоит из следующих шагов:

Шаг 1. Анализ входного вектора, в ходе которого вычисляются количество и значения ненулевых коэффициентов, количество нулей, стоящих перед каждым ненулевым коэффициентов и общее количество нулей.

Шаг 2. Кодирование общего количества ненулевых коэффициентов и количества остаточных единиц.

Шаг 3. Кодирование знаков остаточных единиц.

Шаг 4. Кодирование ненулевых коэффициентов.

Шаг 5. Кодирование общего количества нулей.

Шаг 6. Кодирование серий нулей, предшествующих ненулевым коэффициентам.

Варианты использования векторизации при реализации CALVC

1. Использование MSA для векторного кодирования элементов блока.

Рассмотрим возможность использования MSA для каждого шага алгоритма.

Шаг 1. На данном шаге не рационально использовать векторные инструкции, так как векторные инструкции не подходят для решения задач поиска элементов в массиве.

Шаг 2. Данный шаг состоит только из одной операции поиска в таблице, его реализация на MSA не рациональна.

Шаг 3. Ввиду маленького количества вычислений и простоты реализации данный шаг не нуждается в использовании векторизации.

Шаг 4. Данный шаг требует самого большого количества вычислений (из всех шагов данного алгоритма), а возможность параллельного кодирования коэффициентов позволяет реализовать его на MSA. Однако, реализация MSA затрудняется по следующим двум причинам: а) существу-

ет переменная, которая зависит от кодирования предыдущих коэффициентов; б) наличие 3 if-else операторов, реализация которых на MSA требует использования дополнительных MSA инструкций. Вычисление значения переменной до цикла позволяет решить проблему из пункта «а», а использование дополнительных MSA инструкций позволяет решить проблему из пункта «б».

Шаг 5. Данный шаг состоит только из одной операции поиска в таблице, его реализация на MSA не рациональна.

Шаг 6. Каждый шаг кодирования серии нулей требует информации о количестве оставшихся нулей, а само кодирование происходит путем поиска в таблице. Выбор кода зависит от количества оставшихся нулей и количества нулей в серии, что делает нерациональным кодирование серий нулей инструкциями MSA.

Результаты реализации

В зависимости от вида входного вектора процент использования MSA инструкций может меняться. Рассмотрим несколько видов входных данных:

- Почти все нулевые коэффициенты.
- Равное количество нулевых и ненулевых коэффициентов.
- Все ненулевые коэффициенты.

Количество итераций составляло 1000.

Почти все нулевые коэффициенты

Входная последовательность: 0, 3, 0, 0, 0, 1, 0, 0, 0, -1, 0, -1, 0, 0, 0, 0.

Количество операций MSA: 131000.

Количество обычных операций: 841001.

Процент команд MSA: 13.5%.

Равное количество нулевых и ненулевых коэффициентов

Входная последовательность: 0, -5, 4, 0, 0, -3, 3, 0, 0, 2, 0, -1, 1, 1, 0, 0.

Количество операций MSA: 262000.

Количество обычных операций: 1418001.

Процент команд MSA: 15.6%.

Все ненулевые коэффициенты

Входная последовательность: 12, -10, -8, 6, 6, -5, 4, 4, 3, -3, -3, 2, 2, -1, -1, 1.

Количество операций MSA: 524000.

Количество обычных операций: 1853010.

Процент команд MSA: 28.3%.

Результаты реализации показывают, что чем больше ненулевых коэффициентов, тем больше процент операций MSA, что обусловлено применением векторизации для кодирования ненулевых коэффициентов.

2. Использование MSA для векторного кодирования блоков

При данной реализации на вход алгоритма будут поступать 4 массива по 16 элементов в каждом. Из этих массивов будут составлены 16 векторов по 4 элемента в каждом. Элементы вектора – коэффициенты из массивов, имеющие одинаковый индекс. Поэтому инструкции MSA будут применяться ко всем коэффициентам в массиве с одинаковыми индексами.

Рассмотрим возможность использования MSA для каждого шага алгоритма.

Шаг 1. На данном этапе происходит анализ входных массивов и формируются векторы с ненулевыми коэффициентами, вектор-количество ненулевых коэффициентов в каждом массиве и вектора с количеством нулевых коэффициентов, стоящих перед ненулевыми коэффициентами. Основной задачей данного шага является формирование из неупорядоченных данных MSA векторов для их последующей обработки.

Шаг 2. Ввиду использования таблицы, кодирование общего количества ненулевых коэффициентов и остаточных единиц не нуждается в использовании инструкций MSA.

Шаг 3. Так как для кодирования остаточных единиц используются битовые операции, данные в векторах остаточных единиц не зависят друг от друга, то целесообразно использование инструкций MSA.

Шаг 4. Данный шаг требует большого количества вычислений и хорошо реализуется на MSA. Однако ввиду разного количества ненулевых коэффициентов в каждом из массивов, появляются избыточные операции. Так же на данном шаге присутствуют if-else, что так же вводит дополнительные инструкции MSA.

Шаг 5. Данный шаг состоит только из одной операции поиска в таблице, его реализация на MSA не рациональна.

Шаг 6. В данном шаге используется 2 операции поиска и 1 арифметическая операции, а реализация данного шага на MSA использует дополнительные инструкции, поэтому такая реализация не рациональна.

Результаты реализации

В зависимости от вида входных данных процент использования MSA инструкций может меняться. Рассмотрим вида входных данных:

- Все векторы имеют много нулевых коэффициентов.
- Все векторы имеют мало ненулевых коэффициентов.
- Векторы имеют большую дисперсию количества нулевых и ненулевых коэффициентов.

Количество итераций составляло 1000.

Все векторы имеют много нулевых коэффициентов

Количество операций MSA: 1378216.

Количество обычных операций: 3856815.

Процент команд MSA: 73.7%.

Все векторы имеют мало ненулевых коэффициентов

Количество операций MSA: 2828200.

Количество обычных операций: 3569747.

Процент команд MSA: 55.8%.

Векторы имеют большую дисперсию количества нулевых и ненулевых коэффициентов

Количество операций MSA: 2837282.

Количество обычных операций: 3828787.

Процент команд MSA: 57.4%.

Пример системы для использования отечественного векторного процессора

Вопрос защиты стратегически важных объектов всегда актуален, а обеспечение безопасности водных и прибрежных объектов является сложной задачей, требующей высококачественного решения. В современных условиях это просто необходимо в связи с проблемами терроризма и острой внешнеполитической ситуацией. Автоматизация и упрощение работы гидроакустика позволит в режиме реального времени обнаруживать потенциальные угрозы и своевременно принимать меры по оповещению соответствующих служб. Разрабатываемая система на основе перспективной процессорной архитектуры позволит обеспечивать безопасность прибрежных и водных объектов, таких как мосты, водозаборы, нефтяные платформы и др.

Преимущества системы по сравнению с существующими:

- Высокая скорость обработки и визуализации данных.
- Введение векторного сопроцессора позволит выполнять вычисления над массивами данных в разы быстрее.
- Возможность использования на различных радиусах действия.
- Высокая производительность системы позволяет эффективно и своевременно реагировать на изменение тактической обстановки на любых радиусах действия.

Система состоит из трех модулей:

1) Модуль обработки входных сигналов.
2) Модуль распознавания и классификации целей.
3) Автоматизированное рабочее место (АРМ) оператора.

На вход системы поступает информация об уровнях сигналов в определенном радиусе действия. Эта информация обрабатывается и поступает на систему визуализации, где на представ-

ляется в трехмерном виде. 3D визуализация требует большого количества матричных вычислений, а обработка входных сигналов – классическая задача, решаемая на векторных процессорах, именно поэтому их использование значительно ускорит процедуру визуализации.

По представленным кадрам оператор может сделать вывод о тактической обстановке в радиусе действия системы. Кроме того, информация поступает на вход модуля распознавания и классификации целей, на котором она анализируется, и выводятся подсказки и сообщения о потенциальных угрозах.

В системе будет применены методы трехмерной визуализации, являющиеся принципиально новыми для задач гидроакустического мониторинга. Все вычисления над большими массивами данных будут производиться на векторном сопроцессоре с использование высокоэффективных алгоритмов. Пример работы системы приведён на рисунке 2.

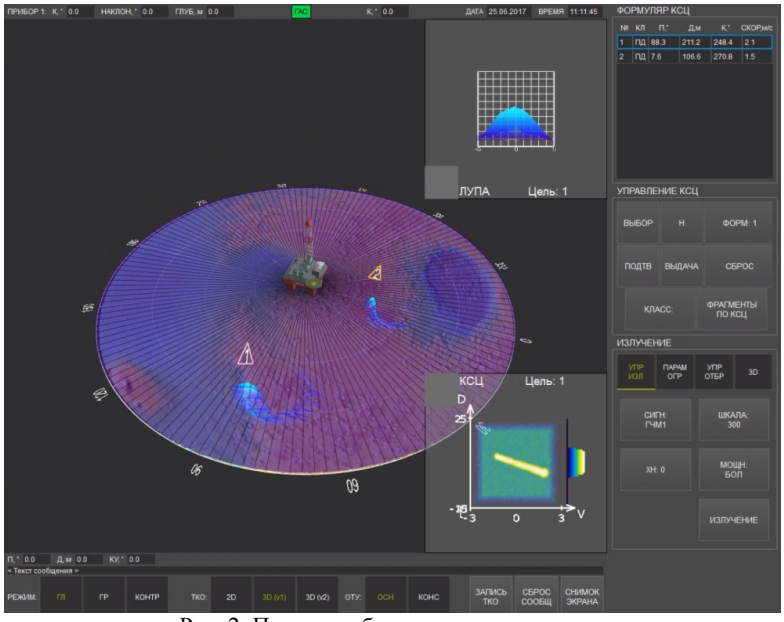


Рис. 2. Пример работы системы.

Заключение

В ходе работы были получены следующие результаты:

1. Перечень наиболее ресурсоёмких вычислительных операций, определенный с помощью профилирования на виртуальной машине QEMU. Это позволило оценить производительность эмулируемой аппаратной платформы с ядром MIPS на исследуемых задачах, а также выбрать алгоритмы обработки изображений и визуализации наиболее эффективные для предложенной архитектуры с точки зрения производительности.
2. Модель архитектуры векторного сопроцессора, включающая модель декодера инструкций вычислительного блока арифметического ядра, модель блока обмена данными с регистровым файлом арифметического ядра, модель регистрового файла арифметического ядра, модель устройства управления передачей данных векторного процессора, модель блока обмена данными с кэш-памятью 1 уровня векторного процессора, модель блока обмена данными с кэш-памятью 2 уровня, модель блока обмена данными с CPU векторного процессора. Это позволило сформировать список предложений для эффективного выполнения исследуемых задач на проектируемой архитектуре векторного сопроцессора.
3. Аппаратные прототипы архитектуры векторного сопроцессора обработки изображений с высокой производительностью и низким энергопотреблением на базе системы остаточных классов, реализованные на FPGA и ASIC. Ожидается, что реализация некоторых операций в системе остаточных классов позволит улучшить параметры производительности и энергопотребления на 10-20% по сравнению с аналогичными реализациями в позиционных системах счисления.

Результаты работы могут быть использованы при построении высокопроизводительных систем обработки изображений, встраиваемых систем анализа многомерных сигналов на основе рекуррентных диаграмм высокого разрешения, при разработке суперкомпьютеров, предназначенных для имитационного моделирования сложных естественных и технических процессов.

ЛИТЕРАТУРА

1. D. Callahan, J. Dongarra, and D. Levine. Vectorizing compilers: A test suite and results. In Proceedings of the 1988 ACM/IEEE Conference on Supercomputing, Supercomputing ’88, pages 98-105, Los Alamitos, CA, USA, 1988. IEEE Computer Society Press.
2. S. Maleki, Y. Gao, M. Garzaran, T. Wong, and D. Padua. An evaluation of vectorizing compilers. In Parallel Architectures and Compilation Techniques (PACT), 2011 International Conference on, pages 372-382, Oct 2011.
3. Виноградов М.А. Современные средства визуализации и обработки двумерных научных данных [Электронный ресурс]. – AMLab, 2002.
4. Ahrens J., Geveci B., Law C. ParaView: An End-User Tool for Large DataVisualization / edited by C.D. Hansen, C.R. Johnson. – Elsevier, 2005. – 984 p.
5. FEI Visualization Sciences Group. Avizo Fire for Core samples and Digital rockphysics [Электронныйресурс]. – 2013. – 4p. – URL:http://www.vsg3d.com/sites/default/files/related/fei_sh_core_sample_digital_rock_physics.pdf.
6. Корякин Ю.А. и др. Корабельная гидроакустическая техника / Ю.А. Корякин – СПб. Наука. – 2004 – 451с.
7. T. Henretty, R. Veras, F. Franchetti, L.-N. Pouchet, J. Ramanujam, and P. Sadayappan. A stencil compiler for short-vector simd architectures. In Proceedings of the 27th International ACM Conference on International Conference on Supercomputing, ICS ’13, pages 13-24, New York, NY, USA, 2013. ACM.
8. D. Nuzman and A. Zaks. Outer-loop vectorization: Revisited for short simd architectures. In Proceedings of the 17th International Conference on Parallel Architectures and Compilation Techniques, PACT ’08, pages 2-11, New York, NY, USA, 2008. ACM.
9. J. Holewinski, R. Ramamurthi, M. Ravishankar, N. Fauzia, L.-N. Pouchet, A. Rountev, and P. Sadayappan. Dynamic tracebased analysis of vectorization potential of applications. In Proceedings of the 33rd ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI ’12, pages 371–382, New York, NY, USA, 2012. ACM.
10. G. C. Evans, S. Abraham, B. Kuhn, and D. A. Padua. Vector seeker: A tool for finding vector potential. In Proceedings of the 2014 Workshop on Programming Models for SIMD/Vector Processing, WPMVP ’14, pages 41–48, New York, NY, USA, 2014. ACM.
11. R. Barik, J. Zhao, and V. Sarkar. Automatic vector instruction selection for dynamic compilation. In Proceedings of the 19th International Conference on Parallel Architectures and Compilation Techniques, PACT ’10, pages 573-574, New York, NY, USA, 2010. ACM.
12. M. Hohenauer, F. Engel, R. Leupers, G. Ascheid, and H. Meyr. A simd optimization framework for retargetable compilers. ACM Trans. Archit. Code Optim., 6(1):2:1–2:27, Apr. 2009.
13. D. S. McFarlin, V. Arbatov, F. Franchetti, and M. Püschel. Automatic simd vectorization of fast fourier transforms for the larrabee and avx instruction sets. In Proceedings of the International Conference on Supercomputing, ICS’11, pages 265–274, New York, NY, USA, 2011. ACM.
14. A. Omondi, B. Premkumar. Residue Number Systems: Theory and Implementation, World Scientific Publishing, 2007, 296 p.
15. P.V.A. Mohan, Residue Number Systems: Algorithms and Architectures, Springer, 2012, 254 p.
16. Н.А. Галанина. Методы и вычислительные устройства цифровой обработки сигналов в системе остаточных классов, диссертация на соискание ученой степени доктора технических наук, Казань, 2010, 268 с.
17. D. Younes and P. Steffan. Efficient image processing application using Residue Number System, Proceedings of the 20th International Conference on Mixed Design of Integrated Circuits and Systems, June 2013. Pp. 468-472.
18. E. Vasalos, D. Bakalis and H.T. Vergos. RNS Assisted Image Filtering and Edge Detection, IEEE 18th International Conference on Digital Signal Processing (DSP), July 2013. Pp. 1–6.
19. N.I. Chervyakov, P.A. Lyakhov, A.S. Ionisyan and M.V. Valueva. High-speed smoothing filter in the Residue Number System, IEEE Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC), July 2016. Pp. 121-126.
20. Azadeh Safari, C.V. Niras and Yinan Kong. Power-performance enhancement of two-dimensional RNS-based DWT image processor using static voltage scaling, Integration, the VLSI journal, vol. 53, March 2016, 2016, pp. 145-156.
21. Jienan Chen, Jianhao Hu. Energy-Efficient Digital Signal Processing via Voltage-Overscaling-Based Residue Number System. IEEE Transactions on Very Large Scale Integration, Vol. 21, No. 7. 2013. Pp. 1322-1332.
22. P. Patronik, K. Berezowski, S.J. Piestrak et. Al. Fast and energy-efficient constant-coefficient FIR filters using residue number system, IEEE International Symposium on Low Power Electronics and Design, 2011. Pp. 385-390.
23. A.S. Molahosseini, S. Sorouri, A.A.E. Zarandi. Research challenges in next-generation residue number system architectures. IEEE 7th International Conference on Computer Science & Education, 2012. Pp. 1658-1661.
24. Бурцев А.А. О возможности оптимизации некоторых функций библиотеки линейной алгебры с помощью векторного сопроцессора// Труды НИИСИ РАН. — 2014. — Т.4, Н.2.— С.5 –15.
25. Borko Furht. Handbook on Multimedia Computing, CRC Press, Inc., Boca Raton, FL, 1998.

D.V. Bogaevskiy, D.I. Kaplun, S.N. Ezhov, A.D. Koshkin (Saint Petersburg Electrotechnical University "LETI", St. Petersburg), **S.I. Aryashev** (Federal State Institution "Scientific Research Institute for System Analysis of the Russian Academy of Sciences", Moscow)

Research of the possibilities of using advanced domestic vector processors in signal processing and multidimensional visualization

In research the minimum set of commands of the vector coprocessor allowing to carry out effectively calculations with material arithmetic in problems of processing of signals and images is allocated. Are made: choice of methods of profiling; preparation of data and modification of the emulator; profiling in various parameters; procedures of the analysis and the check of results allowing to estimate parameters of algorithms for the emulated processor. Use technique in the system of commands of the processor of integer arithmetic of small word length (8, 16, 32 categories) is developed for effective implementation of traditional user tasks, such as coding/decoding of video and audio of data, processing of static images. Results of work can be used at creation of high-performance systems of processing of images, the built-in systems of the analysis of multidimensional signals, when developing the supercomputers intended for imitating modeling of difficult natural and technical processes.

АЛГОРИТМЫ КАЛИБРОВКИ И ДИСКРЕТНОЙ ФИЛЬТРАЦИИ СИГНАЛОВ В АСТРОИНЕРЦИАЛЬНОЙ СИСТЕМЕ ДЛЯ ОПРЕДЕЛЕНИЯ ОРИЕНТАЦИИ МАНЕВРИРУЮЩЕГО СПУТНИКА ЗЕМЛЕОБЗОРА

Е. И. Сомов[1], **С. А. Бутырин**[2], **Т. Е. Сомова**[3]
(Самарский государственный технический университет
Самара, Молодогварлейская ул., 244,
e_somov@mail.ru, butyrinsa@mail.ru, te_somova@mail.ru)

Аннотация

Представляются алгоритмы фильтрации измерений, полетной калибровки и юстировки астроинерциальной системы для определения кинематических параметров ориентации маневрирующего спутника землеобзора.

Ключевые слова: космический аппарат, определение ориентации, астроинерциальная система.

Введение

Рассматривается астроинерциальная система определения углового положения (СОУП) маневрирующего космического аппарата (КА). Данная система содержит инерциальный измерительный модуль (ИИМ) в виде гироскопических измерителей приращений угловых квазикоординат КА и астрономическую систему (АС) на основе звёздных датчиков с широким полем зрения, закрепленных на корпусе спутника. СОУП входит в состав бесплатформенной инерциальной навигационной системы (БИНС), которая решает общую задачу навигации – определения как ориентации, так и местоположения спутника [1 - 4]. Далее считаются известными модели погрешностей ИИМ и АС, понятия и методы обработки информации в БИНС [3 - 6].

В докладе развивается подход к обработке сигналов в астроинерциальной СОУП, предложенный в [7,8] для маневрирующих КА радиоэлектронного и оптико-электронного наблюдения Земли. Проблемы обработки сигналов в СОУП связаны с интегрированием кинематических уравнений по информации только о векторе дискретных приращений угловых квазикоординат, получаемого ИИМ при наличии шумов измерений, с калибровкой (идентификацией и компенсацией вектора дрейфа ИИМ $\mathbf{b}^{\mathrm{g}}$ и погрешности m масштабного коэффициента измерения вектора угловой скорости $\boldsymbol{\omega}$) и с юстировкой [9] – идентификацией и компенсацией погрешности взаимной угловой установки систем координат ИИМ и АС на основе сигналов АС с периодом дискретности $T_o >> T_q$. При решении таких задач многими авторами использовались кинематические параметры в виде кватерниона $\boldsymbol{\Lambda} = (\lambda_0, \boldsymbol{\lambda})$, матрицы ориентации $\mathbf{C}$, вектора Эйлера $\boldsymbol{\phi} = \mathbf{e}\,\Phi$, вектора конечного поворота $\boldsymbol{\Phi} = 2\mathbf{e}\,\mathrm{tg}(\Phi/2)$ и др. При этом для медленных движений КА с практически постоянным ортом Эйлера $\mathbf{e}$ интегрирование кинематических соотношений для вектора Эйлера и получение значений кватерниона $\boldsymbol{\Lambda}_k \equiv \boldsymbol{\Lambda}(t_k)$ выполнялось по схеме

$$\delta\boldsymbol{\phi}_k = \delta\Phi_k \mathbf{e}_k \equiv \mathbf{i}_k^{\omega} = \int_{t_k}^{t_{k+1}} \boldsymbol{\omega}(\tau) d\tau \equiv \mathbf{Int}(t_k, T_o, \boldsymbol{\omega}(t)); \quad \boldsymbol{\phi}_k + \delta\boldsymbol{\phi}_k = \boldsymbol{\phi}_{k+1} \Rightarrow \mathbf{C}_{k+1} \Rightarrow \boldsymbol{\Lambda}_{k+1},$$

где $t_{k+1} = t_k + T_o$, $k \in \mathrm{N}_0 \equiv [0,1,2,...)$.

Угловое движение маневрирующего КА землеобзора выполняется на последовательности временных интервалов маршрутов сканирования с модулем вектора угловой скорости $\omega \le \omega_1^{\mathrm{m}} = 1$ град/с и быстрых поворотных маневров (ПМ) с переменным направлением вектора $\boldsymbol{\omega}$ и модулем $\omega \le \omega^{\mathrm{m}} = 3$ град/сек. Для вектора Эйлера дифференциальное уравнение имеет вид $\dot{\boldsymbol{\phi}} = \boldsymbol{\omega} + \frac{1}{2}\boldsymbol{\phi} \times \boldsymbol{\omega} + f(\theta)\boldsymbol{\phi} \times (\boldsymbol{\phi} \times \boldsymbol{\omega})$, где скалярная функция $f(\theta) = (1 - \theta\sin\theta/(2(1-\cos\theta)))/\theta^2$ [10], при малых значениях угла θ функция $f(\theta) \approx 1/12$. На

[1] Кандидат технических наук, начальник отдела Навигации, наведения и управления движением.
[2] Кандидат технических наук, начальник лаборатории Моделирования систем управления.
[3] Младший научный сотрудник лаборатории Моделирования систем управления.

основе этого уравнения нетрудно понять физический смысл возникающего конинга (coning) – известного эффекта некоммутативного поворота, обусловленного изменением именно направления вектора $\boldsymbol{\omega}$ в процессе выполнения поворотного маневра [10 - 12].

В статье рассматривается классическая схема СОУП в составе БИНС, где основным измерителем является ИИМ, а сигналы АС с периодом T_o применяются для его калибровки и юстировки. При этом кластер гироскопических датчиков угловой скорости в составе ИММ имеет выходной вектор приращений угловых квазикоординат – интегралов от проекций измеренного вектора $\boldsymbol{\omega}$ на периоде дискретности T_q, кратного периоду T_o. Здесь применяются оригинальные авторские приёмы полиномиальной аппроксимации и интерполяции, выполняется численное интегрирование кинематического уравнения для вектора модифицированных параметров Родрига (МПР) $\boldsymbol{\sigma} = \mathbf{e}\,\mathrm{tg}(\Phi/4)$ [13], который связан с кватернионом $\boldsymbol{\Lambda}$ прямыми $\boldsymbol{\sigma} = \boldsymbol{\lambda}/(1+\lambda_0)$ ($\boldsymbol{\Lambda} \Rightarrow \boldsymbol{\sigma}$) и обратными $\boldsymbol{\lambda} = 2\boldsymbol{\sigma}/(1+\sigma^2)$; $\lambda_0 = (1-\sigma^2)/(1+\sigma^2)$ ($\boldsymbol{\sigma} \Rightarrow \boldsymbol{\Lambda}$) соотношениями. Кинематические прямые и обратные уравнения для вектора МПР имеют вид

$$\dot{\boldsymbol{\sigma}} = \mathbf{F}^{\sigma}(\boldsymbol{\sigma},\boldsymbol{\omega}) \equiv \tfrac{1}{4}(1-\sigma^2)\boldsymbol{\omega} + \tfrac{1}{2}\boldsymbol{\sigma}\times\boldsymbol{\omega} + \tfrac{1}{2}\boldsymbol{\sigma}\langle\boldsymbol{\sigma},\boldsymbol{\omega}\rangle\,; \quad \boldsymbol{\omega} = 4[(1-\sigma^2)\dot{\boldsymbol{\sigma}} - 2(\boldsymbol{\sigma}\times\dot{\boldsymbol{\sigma}}) + 2\boldsymbol{\sigma}\langle\dot{\boldsymbol{\sigma}},\boldsymbol{\sigma}\rangle]/(1+\sigma^2)^2$$

при стандартном обозначении $\langle \mathbf{a},\mathbf{b}\rangle$ скалярного произведения векторов $\mathbf{a}$ и $\mathbf{b}$.

Статья посвящается светлой памяти профессора Андрея Евгеньевича Барабанова.

Постановка задачи и подход к решению

Аналогично [7,8] вводятся инерциальный базис $\mathbf{I}$; базис $\mathbf{B}$, связанный с корпусом КА; базис $\mathbf{S}$ сенсора (телескопа, радиолокационной антенны); виртуальный базис АС $\mathbf{A}$, вычисляемый на основе информации от кластера звёздных датчиков, и виртуальный базис ИИМ $\mathbf{G}$, вычисляемый на основе информации от датчиков угловых скоростей с выходным вектором приращений угловых квазикоординат. Будем для простоты считать, что базисы $\mathbf{B}$ и $\mathbf{S}$ совпадают, известна информация о взаимном угловом положении базисов $\mathbf{A}$ и $\mathbf{S}$, полученная одним из известных методов, в частности для КА оптико-электронного землеобзора см. [14,15]. Пусть исходная измерительная информация обрабатывается в ИИМ с частотой ≈ 3 кГц и в результате от ИИМ поступают измеренные значения вектора приращений квазикоординат $\mathbf{i}^{\mathrm{g}\omega}_{\mathrm{m}\,s}, s \in \mathrm{N}_0$ с периодом $T_q << T_o$, а от АС – измеренные значения кватерниона $\boldsymbol{\Lambda}^{\mathrm{a}}_{\mathrm{m}\,k}$ с периодом T_o:

$$\mathbf{i}^{\mathrm{g}\omega}_{\mathrm{m}\,s} = \mathbf{Int}(t_s, T_q, \boldsymbol{\omega}^{\,\mathrm{g}}_{\mathrm{m}}(t)) + \boldsymbol{\delta}^{\mathrm{n}}_s\,; \; \boldsymbol{\omega}^{\,\mathrm{g}}_{\mathrm{m}}(t) \equiv (1+m)\mathbf{S}^{\Delta}(\boldsymbol{\omega}(t) + \mathbf{b}^{\mathrm{g}})\,; \;\; \boldsymbol{\Lambda}^{\mathrm{a}}_{\mathrm{m}\,k} = \boldsymbol{\Lambda}_k \circ \boldsymbol{\Lambda}^{\mathrm{n}}_k\,. \qquad (1)$$

Здесь $\boldsymbol{\omega}^{\,\mathrm{g}}_{\mathrm{m}}(t)$ представляет измеряемый вектор угловой скорости КА в базисе $\mathbf{G}$ с учетом неизвестных малых и медленных вариаций вектора $\mathbf{b}^{\mathrm{g}} = \mathbf{b}^{\mathrm{g}}(t)$ дрейфа ИИМ по угловой скорости; при обозначениях $C_\alpha = \cos\alpha$, $S_\alpha = \sin\alpha$ ортогональной матрице $\mathbf{S}^{\Delta}(t)$ погрешности угловой установки систем координат ИИМ и АС соответствует кватернион $\boldsymbol{\Lambda}^{\Delta} = (C_{\varphi^{\Delta}/2}, \mathbf{e}^{\Delta} S_{\varphi^{\Delta}/2})$ с неизвестными ортом $\mathbf{e}^{\Delta} = \{\mathrm{e}^{\Delta}_x, \mathrm{e}^{\Delta}_y, \mathrm{e}^{\Delta}_z\}$ и углом Эйлера φ^{Δ}; скалярная функция $m = m(t)$ представляет неизвестную медленную вариацию масштабного коэффициента ИИМ, учтены центрированные дискретные гауссовские шумы $\boldsymbol{\delta}^{\mathrm{n}}_s$ со среднеквадратичным отклонением (СКО) σ^{b} и $\boldsymbol{\Lambda}^{\mathrm{n}}_k$ с СКО σ^{a} в выходных сигналах ИИМ и АС соответственно. Зависимость $\sigma^{\mathrm{a}}(\omega)$ является монотонно возрастающей, причем $\sigma^{\mathrm{a}}(0)/\sigma^{\mathrm{a}}(\omega^{\mathrm{m}}_1) \approx 1/3$. Будем считать, что матрица $\mathbf{S}^{\Delta}(t)$ близка к единичной $\mathbf{I}_3$ с представлением $\mathbf{S}^{\Delta} \cong \mathbf{I}_3 + [\boldsymbol{\Delta}\times]$, где вектор «скосов» $\boldsymbol{\Delta} = \{\Delta_x, \Delta_y, \Delta_z\}$ систем координат ИИМ и АС имеет компоненты $\Delta_i = \mathrm{e}^{\Delta}_i \varphi^{\Delta}$, $i = x, y, z$, и что вариация масштабного коэффициента мала, например $|m(t)| \le 0.01$, когда справедливо соотношение $1 - m^2 \cong 1$.

Задача состоит в разработке алгоритмов калибровки и юстировки СОУП по сигналам АС с периодом T_o с получением оценок $\hat{\mathbf{b}}^{\mathrm{g}}_k$, $\hat{m}_k$ и $\hat{\mathbf{S}}^{\Delta}_k$, дискретной рекуррентной фильтрации изме-

рений ИИМ с периодом T_q, а также алгоритмов цифровой фильтрации выходных сигналов СОУП с периодом $T_p = t_{l+1} - t_l$, кратном периоду T_o, в общем случае с фиксированным запаздыванием T_d относительно моментов времени t_k, с получением дискретных значений кватерниона $\mathbf{\Lambda}_l^{\mathrm{f}}$ и вектора угловой скорости $\boldsymbol{\omega}_l^{\mathrm{f}}$, $l \in \mathrm{N}_0$, которые используются для управления ориентацией маневрирующего КА землеобзора, при наличии быстрых ПМ спутника с переменным направлением вектора его угловой скорости на ограниченных временных интервалах.

Подход к решению задачи. В модели (1) вектор дрейфа $\mathbf{b}^{\mathrm{g}}(t)$ представляет аддитивное возмущение, а матрица $\mathbf{S}^{\Delta}(t)$ и вариация масштабного коэффициента $m(t)$ – мультипликативные возмущения, идентификация которых является наиболее сложна. В известных методах решения таких задач для корректируемых БИНС [9,13] используются нестационарные фильтры Калмана, которые «работают» в малой окрестности тестовых угловых движений КА, которые необходимы для обеспечения условий наблюдаемости. Применяемый авторский подход основывается на дискретной фильтрации доступных измерений и следующих положениях [7,8]:

- для калибровки ИИМ с периодом T_o в отношении дрейфа $\mathbf{b}^{\mathrm{g}}$ применяется дискретный стационарный нелинейный фильтр Луенбергера с параметрами, при которых обеспечиваются показатели качества оценки $\hat{\mathbf{b}}^{\mathrm{g}}$, близкие к аналогичным показателям соответствующего фильтра Калмана, причём такая калибровка выполняется на временных интервалах, где модуль угловой скорости КА не превышает 1 град/с и измерения АС являются вполне точными;
- при ПМ спутника с модулем угловой скорости от 1 град/с до 3 град/с на временных интервалах длительностью до 120 с астрономическая коррекция СОУП временно отключается и оценка ориентации КА выполняется с использованием прогноза вариации вектора дрейфа ИИМ, который формируется на основе анализа тренда оценки $\hat{\mathbf{b}}^{\mathrm{g}}$ в «скользящем окне» на предшествующем интервале работы СОУП с астрономической коррекцией;
- юстировка СОУП (оценка $\hat{\mathbf{S}}^{\Delta}$) и её калибровка в отношении вариации масштабного коэффициента с определением оценки $\hat{m}$ выполняются off-line на основе сопоставления значений вектора $\boldsymbol{\omega}$, которые оцениваются автономно и одновременно по сигналам ИИМ (значений вектора $\hat{\boldsymbol{\omega}}^{\mathrm{g}}$) и астрономической системы (значений вектора $\hat{\boldsymbol{\omega}}^{\mathrm{a}}$).

Отметим, что в предлагаемом подходе полётная калибровка и юстировка СОУП выполняются на фоне решения целевых задач маневрирующего спутника, без каких-либо тестовых пространственных угловых движений КА.

Дискретная фильтрация измерений и компенсация дрейфа ИИМ

Будем предусматривать формирование оценок $\hat{\mathbf{b}}_k^{\mathrm{g}}$, $\hat{\mathbf{S}}_k^{\Delta}$ и $\hat{m}_k$, постоянных на периоде T_o, когда оценка $\hat{\mathbf{b}}_k^{\mathrm{g}}$ обновляется on-line (в каждый момент времени t_k), а оценки $\hat{\mathbf{S}}_k^{\Delta}$ и $\hat{m}_k$ регулярно формируются off-line, т.е. на основе обработки доступной измерительной информации, накопленной на длительных временных интервалах. При дискретной фильтрации измерений вектора приращений квазикоординат с целью подавления дискретного шума ИИМ впервые применяется оригинальная двухпроходная технология фильтрации – сочетание аппроксимации измеренных данных $\mathbf{i}_s^{\mathrm{g}\omega}$ полиномом 3-го порядка $\widetilde{\mathbf{i}}_k^{\mathrm{g}\omega}(\tau)$ в скользящем окне 9 измерений по методу наименьших квадратов (МНК) и сплайновой интерполяции по центрам двух смежных скользящих окон полиномом 5-го порядка $\hat{\mathbf{i}}_k^{\mathrm{g}\omega}(\tau)$ для локального времени $\tau = t - k\,T_o \in [0, T_o]$.

Разработанная технология иллюстрируется схемой на рис. 1, где синими «звездочками» отмечены погрешности $\delta \mathrm{i}_s^{\mathrm{g}\omega}$ дискретного измерения скалярной квазикоординаты ИИМ в моменты времени t_s (здесь указаны значения индекса s), зелеными пунктирными линиями выделены

полиномы $\delta\tilde{\mathrm{i}}^{\mathrm{g}\omega}(\tau)$ 3-го порядка, которые отражают погрешности аппроксимации по МНК в скользящем окне 9 измерений, и бордовыми линиями представлены гладко сопряженные полиномы $\delta\hat{\mathrm{i}}^{\mathrm{g}\omega}(\tau)$ 5-го порядка, которые соответствуют погрешностям процедуры интерполяции по центрам двух смежных скользящих окон. В нижней части этого рисунка приведена погрешность $\delta\omega^{\mathrm{g}}(\tau)$ оценки $\hat{\omega}^{\mathrm{g}}(\tau)$ угловой скорости в базисе $\mathbf{G}$. При этом оценка $\hat{\omega}^{\mathrm{g}}(\tau)$ получается явным дифференцированием и строго согласована с оценкой $\hat{\mathrm{i}}^{\mathrm{g}\omega}(\tau)$ квазикоординаты ИИМ.

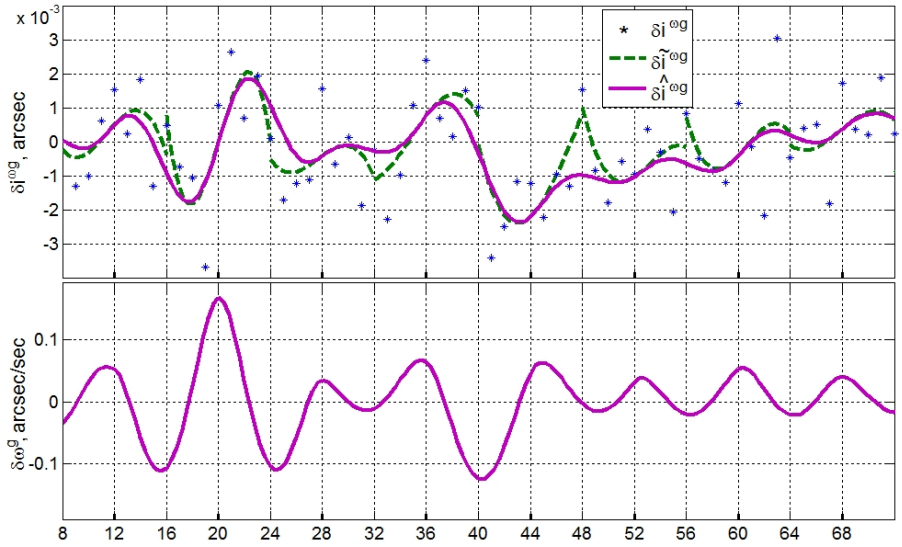


Рис. 1. Погрешности двухпроходной технологии фильтрации измерений ИММ

При компенсации влияния вектора дрейфа $\mathbf{b}^{\mathrm{g}}$ и погрешности масштабного коэффициента m в базисе $\mathbf{G}$ получается оценка векторного полинома (сплайна) приращений квазикоординат в виде $\hat{\mathbf{i}}_k^{\mathrm{g}\omega}(\tau) = (1-\hat{m}_k)(\tilde{\mathbf{i}}_k^{\mathrm{g}\omega}(\tau) - \hat{\mathbf{b}}_k^{\mathrm{g}}\tau)$. В базисе $\mathbf{A}$ астрономической системы оценка непрерывного вектора приращений квазикоординат $\hat{\mathbf{i}}_k^{\omega}(\tau)$ вычисляется по формуле

$$\hat{\mathbf{i}}_k^{\omega}(\tau) = (\hat{\mathbf{S}}_k^{\Delta})^{\mathrm{t}}\hat{\mathbf{i}}_k^{\mathrm{g}\omega}(\tau) = (\hat{\mathbf{S}}_k^{\Delta})^{\mathrm{t}}(1-\hat{m}_k)(\tilde{\mathbf{i}}_k^{\mathrm{g}\omega}(\tau) - \hat{\mathbf{b}}_k^{\mathrm{g}}\tau), \qquad (2)$$

а непрерывная оценка $\hat{\boldsymbol{\omega}}_k(\tau)$ значений вектора угловой скорости $\boldsymbol{\omega}(t)$ на k-ом интервале времени $t \in [t_k, t_{k+1}]$ с получается аналитическим дифференцированием этого векторного сплайна, что обеспечивает полное соответствие векторных полиномов $\hat{\mathbf{i}}_k^{\omega}(\tau)$ и $\hat{\boldsymbol{\omega}}_k(\tau)$.

Идентификация дрейфа $\mathbf{b}^{\mathrm{g}}$ ИИМ выполняется с применением дискретного наблюдателя Луенбергера при периоде дискретности T_o. Для определения структуры этого наблюдателя рассмотрим сначала упрошенную линеаризованную модель ИИМ, где $\mathbf{S}^{\Delta} \equiv \mathbf{I}_3$, $m(t) \equiv 0$, орт $\mathbf{e}_k(\tau)$ вектора МПР $\boldsymbol{\sigma}_k(\tau) = \mathbf{e}_k(\tau)\,\mathrm{tg}(\theta_k(\tau)/4)$ неизменен в базисе $\mathbf{I}$, т.е. $\mathbf{e}_k(\tau) \cong \mathbf{e}_k$ на k-ом временном интервале с локальным временем $\tau = t - k\,T_o \in [0, T_o]$, а вектор дрейфа $\mathbf{b}^{\mathrm{g}}(\tau)$ считается постоянным на этом временном интервале, т.е. $d\,\mathbf{b}_k^{\mathrm{g}}(\tau)/d\,\tau \equiv \dot{\mathbf{b}}_k^{\mathrm{g}}(\tau) = \mathbf{0} \quad \forall \tau \in [0, T_o]$. Если при этом шум измерений ИИМ и АС не учитывается, то дискретное изменение значений вектора $\mathbf{p}_k = 4\,\boldsymbol{\sigma}_k \approx \boldsymbol{\phi}_k$ с периодом T_o описывается соотношением $\mathbf{p}_{k+1} = \mathbf{p}_k + \mathbf{i}_k^{\omega} + \mathbf{b}_k^{\mathrm{g}}\,T_o$. При отсутствии шума измеряемые значения $\mathbf{p}_k^{\mathrm{a}} = \mathbf{p}_k$ вектора $\mathbf{p}_k$ аналитически определяются по соотношениям $\boldsymbol{\Lambda} \Rightarrow \boldsymbol{\sigma}$ и $\mathbf{p} = 4\,\boldsymbol{\sigma}$ на основе измеряемых дискретных сигналов АС в виде кватерниона $\boldsymbol{\Lambda}_{mk}^{\mathrm{a}} = \boldsymbol{\Lambda}_k$. Здесь отсутствует какая-либо измеряемая информация о значении вектора $\mathbf{b}_k^{\mathrm{g}}$, поэтому решается локальная задача синтеза структуры асимптотически устойчивого линейного

дискретного наблюдателя Луенбергера [16] для получения дискретных оценок $\hat{\mathbf{p}}_k$ и $\hat{\mathbf{b}}^{\mathrm{g}}_k$ векторов $\mathbf{p}_k$ и $\mathbf{b}^{\mathrm{g}}_k$ соответственно. При обозначениях $\{\cdot\}$ и $[\cdot]$ для столбца и строки, векторов состояния $\mathbf{x}_k = \{\mathbf{p}_k, \mathbf{b}^{\mathrm{g}}_k\}$, входа $\mathbf{u}_k = \mathbf{i}^{\omega}_k$ и измеряемого выхода $\mathbf{y}_k = \mathbf{p}_k$ модель дискретной системы представляется в стандартной форме

$$\begin{aligned} &\mathbf{x}_{k+1} = \mathbf{A}_d\,\mathbf{x}_k + \mathbf{B}_d\,\mathbf{u}_k; \\ &\mathbf{y}_k = \mathbf{C}_d\,\mathbf{x}_k, \end{aligned} \qquad \mathbf{A}_d = \begin{bmatrix} \mathbf{I}_3 & T_o\mathbf{I}_3 \\ \mathbf{0} & \mathbf{I}_3 \end{bmatrix}; \qquad \begin{aligned} &\mathbf{B}_d = \{\mathbf{I}_3, \mathbf{0}\}; \\ &\mathbf{C}_d = [\mathbf{I}_3, \mathbf{0}]. \end{aligned} \tag{3}$$

При символе транспонирования $(\cdot)^{\mathrm{t}}$ матрица наблюдаемости $\mathbf{L}^s_d \equiv [\mathbf{C}^{\mathrm{t}}_d, \mathbf{A}^{\mathrm{t}}_d\mathbf{C}^{\mathrm{t}}_d]$ этой системы имеет $\mathrm{rang}(\mathbf{L}^s_d) = 6$, так как $\det(\mathbf{L}^s_d) = T_o^3$, поэтому система (3) полностью наблюдаема [16] и структура линейного дискретного наблюдателя (фильтра) Луенбергера для получения оценки $\hat{\mathbf{x}}_k = \{\hat{\mathbf{p}}_k, \hat{\mathbf{b}}^{\mathrm{g}}_k\}$ вектора $\mathbf{x}_k$ имеет вид

$$\hat{\mathbf{x}}_{k+1} = \mathbf{A}_d\,\hat{\mathbf{x}}_k + \mathbf{B}_d\,\mathbf{u}_k + \mathbf{G}_d(\mathbf{y}_k - \mathbf{C}_d\hat{\mathbf{x}}_k)\,. \tag{4}$$

Здесь постоянная матрица $\mathbf{G}_d$ назначается из условия асимптотической устойчивости тривиального решения $\boldsymbol{\varepsilon}_k \equiv \mathbf{x}_k - \hat{\mathbf{x}}_k = \mathbf{0}$ дискретной системы

$$\boldsymbol{\varepsilon}_{k+1} = \mathbf{A}^o_d\boldsymbol{\varepsilon}_k \equiv (\mathbf{A}_d - \mathbf{G}_d\mathbf{C}_d)\,\boldsymbol{\varepsilon}_k\,, \tag{5}$$

что соответствует принадлежности спектра матрицы $\mathbf{A}^o_d$ внутренности единичного круга, т.е. условию $|z_p| < 1 \quad \forall p = 1 \div 6$, где z_p – корни уравнения $\det(z\mathbf{I}_6 - \mathbf{A}^o_d) = 0$. Рационально выбрать структуру матрицы $\mathbf{G}_d$ в виде $\mathbf{G}_d = \{g^o_1\mathbf{I}_3, g^o_2\mathbf{I}_3\}$, где скалярные коэффициенты g^o_1 и g^o_2 определяются аналитически по каноническим представлениям линейных дискретных систем управления [17] с обеспечением одинаковой степени устойчивости системы (5) по каналам идентификации, т.е. с обеспечением одинаковых показателей сходимости оценок компонентов векторов $\mathbf{p}_k$ и $\mathbf{b}^{\mathrm{g}}_k$. В этом случае характеристическое уравнение каждого канала идентификации имеет вид $z^2 + a^d_1 z + a^d_9 = 0$, где $a^d_1 = -(2 - g^o_1)$ и $a^d_0 = 1 - (g^o_1 - T_o g^o_2)$.

При желаемом спектре канала в виде $z^*_{1,2} = \exp(\nu T_{\mathrm{o}}(-\xi \pm j\sqrt{1-\xi^2}))$, ($j = \sqrt{-1}$), коэффициенты характеристического уравнения вычисляются по формулам Виета $a^{*d}_0 = z^*_1 z^*_2 = \exp(-2\xi\nu T_0)$; $a^{*d}_1 = -(z^*_1 + z^*_2) = -2\exp(-\xi\nu T_0)\cos(\nu T_0\sqrt{1-\xi^2})$. Это позволяет в явном виде определить коэффициенты дискретного наблюдателя Луенбергера g^o_1 и g^o_2 из условий $a^d_1 = a^{*d}_1$ и $a^d_0 = a^{*d}_0$.

В итоге структура линейного дискретного фильтра Луенбергера для упрощенной линейной модели (4) представляется в явном виде

$$\hat{\mathbf{p}}_{k+1} = \hat{\mathbf{p}}_k + \mathbf{i}^{\omega}_k + g^o_1\,\Delta\mathbf{p}_k; \quad \hat{\mathbf{b}}^{\mathrm{g}}_{k+1} = \hat{\mathbf{b}}^{\mathrm{g}}_k + g^o_2\,\Delta\mathbf{p}_k; \quad \Delta\mathbf{p}_{k+1} \equiv \mathbf{p}^{\mathrm{a}}_{k+1} - \hat{\mathbf{p}}_{k+1}, \tag{6}$$

где постоянные коэффициенты g^o_1 и g^o_2 вычисляются по явным аналитическим соотношениям.

При прямом учете шумов измерений АС наблюдатель типа (6), где вектор $\mathbf{p}^{\mathrm{a}}_k$ вычисляется по измерениям АС в виде $\boldsymbol{\Lambda}^{\mathrm{a}}_{\mathrm{m}\,k} = \boldsymbol{\Lambda}_k \circ \boldsymbol{\Lambda}^{\mathrm{n}}_k$ с кватернионом $\boldsymbol{\Lambda}^{\mathrm{n}}_k$ дискретного шума, может представляться также в форме дискретного фильтра Калмана. При этом только для установившегося режима определяются постоянные матрицы ковариации и коэффициентов фильтра Калмана. Как указано выше, для маневрирующих КА землеобзора при выполнении маршрутов сканирующей съемки модуль ω вектора угловой скорости $\boldsymbol{\omega}$ может существенно изменяться, $\omega \le \omega^{\mathrm{m}}_1 = 1$ град/с, причем зависимость СКО $\sigma^{\mathrm{a}}(\omega)$ погрешности АС является монотонно возрастающей, $\sigma^{\mathrm{a}}(0)/\sigma^{\mathrm{a}}(\omega^{\mathrm{m}}_1) \approx 1/3$. Поэтому стремление к оптимальной по МНК идентификации вектора $\mathbf{b}^{\mathrm{g}}_k$ дрейфа ИИМ с помощью фильтра Калмана неминуемо приводит к необходимости численного решения нелинейного матричного уравнения Риккати в реальном масштабе времени. Поэтому здесь применяется наблюдатель дрейфа ИИМ в форме фильтра Луенбергера и его

постоянные параметры назначаются из условий обеспечения показателей качества оценки $\hat{\mathbf{b}}^{\mathrm{g}}$, близких к аналогичным показателям фильтра Калмана в установившемся режиме при значении модуля вектора угловой скорости $\omega = \omega_1^{\mathrm{m}}$.

Возвратимся к задаче идентификации дрейфа ИИМ $\mathbf{b}^g$ с помощью фильтра Луенбергера в СОУП маневрирующего КА, когда орт Эйлера $\mathbf{e}_k(\tau)$ вектора МПР $\boldsymbol{\sigma}_k(\tau)$ существенно изменяет свое направление в базисе $\mathbf{I}$ на k-ом временном интервале с локальным временем $\tau = t - k\,T_o \in [0, T_o]$. Оценка изменения углового положения КА на таком интервале времени достигается численным интегрированием векторного дифференциального уравнения

$$\dot{\hat{\boldsymbol{\sigma}}}_k(\tau) = \mathbf{F}^{\sigma}(\hat{\boldsymbol{\sigma}}_k(\tau), \hat{\boldsymbol{\omega}}_k(\tau)) = \tfrac{1}{4}(1-\hat{\sigma}_k^2)\hat{\boldsymbol{\omega}}_k + \tfrac{1}{2}\hat{\boldsymbol{\sigma}}_k \times \hat{\boldsymbol{\omega}}_k + \tfrac{1}{2}\hat{\boldsymbol{\sigma}}_k \langle \hat{\boldsymbol{\sigma}}_k, \hat{\boldsymbol{\omega}}_k \rangle \qquad (7)$$

методом ODE45 [18,19] с формированием оценки вектора МПР $\hat{\boldsymbol{\sigma}}_k(\tau)$. Вычислительный дрейф такой процедуры интегрирования был детально исследован в [7], например при коническом движении вектора угловой скорости с модулем 0.5 град/с и шаге интегрирования 0,015625 с вычислительный дрейф не превосходит 0.00025 угл. сек за время 720 с. Начальные условия для уравнения (7) формируется непосредственно по сигналам АС только в момент времени t_0 (при включении СОУП), в остальных случаях начальные условия такого интегрирования на очередном k-ом временном интервале определятся по сигналам дискретного фильтра Луенбергера.

Опуская индекс k временного интервала, приведем соотношения для вычисления непрерывных погрешностей при получении оценок углового положения и угловой скорости КА. Оценки кинематических параметров определяются кватернионом $\hat{\boldsymbol{\Lambda}}(\tau)$ и вектором $\hat{\boldsymbol{\omega}}(\tau)$. В результате интегрирования (7) получаются значения $\hat{\boldsymbol{\sigma}}(\tau)$ векторной функции МПР, на основе которых с помощью преобразования $\hat{\boldsymbol{\sigma}}(\tau) \Rightarrow \hat{\boldsymbol{\Lambda}}(\tau)$ формируется оценка $\hat{\boldsymbol{\Lambda}}(\tau)$ изменения кватерниона. Погрешности оценки ориентации определяются кватернионом $\mathbf{E} = \tilde{\boldsymbol{\Lambda}} \circ \hat{\boldsymbol{\Lambda}} \equiv (e_0, \mathbf{e})$ с вектором $\mathbf{e} = \{e_i\}$, матрицей погрешности $\mathbf{C}^{\mathrm{e}} = \mathbf{I}_3 - 2[\mathbf{e}\times]\mathbf{Q}_{\mathrm{e}}^{\mathrm{t}}$, где $\mathbf{Q}_{\mathrm{e}} = \mathbf{I}_3 e_0 + [\mathbf{e}\times]$, вектором МПР $\boldsymbol{\sigma}^{\mathrm{e}} = \mathbf{e}/(1+e_0)$ и столбцом угловых ошибок $\delta\boldsymbol{\phi} = \{\delta\phi_i\} = 4\boldsymbol{\sigma}^{\mathrm{e}}$. Непрерывная погрешность оценки вектора угловой скорости $\delta\boldsymbol{\omega}(\tau)$ определяется соотношением $\delta\boldsymbol{\omega}(\tau) = \hat{\boldsymbol{\omega}}(\tau) - \mathbf{C}^{\mathrm{e}}\boldsymbol{\omega}(\tau)$. На рис. 2 бордовыми линиями схематично представлены непрерывные погрешности при оценке углового положения $\delta\phi$ и угловой скорости $\delta\omega$ КА в базисе $\mathbf{A}$ астрономической системы.

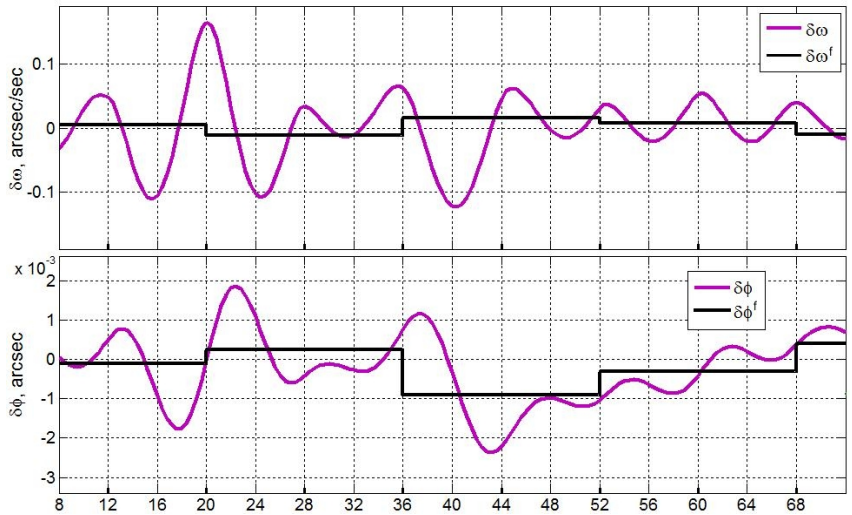


Рис. 2. Погрешности непрерывной оценки и цифровой фильтрации выходных сигналов СОУП

Пусть в момент времени $t = t_k$ имеются оценка матрицы $\hat{\mathbf{S}}_k^{\Delta}$ и погрешности масштабного коэффициента $\hat{m}_k$ измерений вектора угловой скорости, постоянные на периоде T_o (методика

получения таких оценок off-line представлена ниже), оценка кватерниона $\hat{\mathbf{\Lambda}}_k$ и вектора дрейфа $\hat{\mathbf{b}}_k^{\mathrm{g}}$ ИИМ, измеренное АС фактическое угловое положение КА в виде кватерниона $\mathbf{\Lambda}_{\mathrm{m}\,k}^{\mathrm{a}}$, а также на предыдущем шаге сформированы кватернион $\Delta\mathbf{P}_k(g_1^o,\mathbf{Q}_k)$ и вектор $\Delta\mathbf{p}_k(g_2^o,\mathbf{Q}_k)$, где кватернион $\mathbf{Q}_k \equiv (q_{0k},\mathbf{q}_k) \equiv (C_{\varphi_k/2},\mathbf{e}_k^q S_{\varphi_k/2}) \equiv \mathbf{Q}_k(\mathbf{e}_k^q,\varphi_k) = \widetilde{\mathbf{\Lambda}}_{\mathrm{m}\,k}^{\mathrm{a}} \circ \hat{\mathbf{\Lambda}}_k$. В этот же момент времени t_k с помощью преобразования $\hat{\mathbf{\Lambda}}_k \Rightarrow \hat{\boldsymbol{\sigma}}_k$ определяется начальное условие $\hat{\boldsymbol{\sigma}}_k(0) \equiv \hat{\boldsymbol{\sigma}}_k$ численного интегрирования уравнения (7) для получения оценки вектора МПР $\hat{\boldsymbol{\sigma}}_k(\tau)$ на k-ом интервале. Такое интегрирование выполняется для указанной выше последовательности гладко сопряженных векторных сплайнов $\hat{\boldsymbol{\omega}}_k(\tau)$. Здесь рационально применять параллельные вычисления для аппроксимации и интерполяции при получении в базисе $\mathbf{G}$ векторного сплайна $\hat{\mathbf{i}}_k^{\mathrm{g}\omega}(\tau)$, его представления $\hat{\mathbf{i}}_k^{\omega}(\tau)$ (2) в базисе $\mathbf{A}$ астрономической системы с аналитическим формированием оценки $\hat{\boldsymbol{\omega}}_k(\tau)$ и собственно численного интегрирования (7). В момент локального времени $\tau = T_o$ в результате интегрирования получается значение вектора МПР $\hat{\boldsymbol{\sigma}}_{k+1} = \hat{\boldsymbol{\sigma}}_k(T_o)$, по которому с помощью преобразования $\hat{\boldsymbol{\sigma}}_{k+1} \Rightarrow \hat{\mathbf{R}}_{k+1}$ определяется значение кватерниона $\hat{\mathbf{R}}_{k+1}$. Разработанный нелинейный дискретный наблюдатель Луенбергера, где используются численное интегрирование (7) в сочетании с кватернионными и векторными соотношениями, в конечном итоге имеет вид

$$\begin{aligned} &\hat{\mathbf{\Lambda}}_{k+1} = \hat{\mathbf{R}}_k \circ \Delta\mathbf{P}_k(g_1^o,\mathbf{Q}_k); && \hat{\mathbf{b}}_{k+1}^{\mathrm{g}} = \hat{\mathbf{b}}_k^{\mathrm{g}} + \Delta\mathbf{p}_k(g_2^o,\mathbf{Q}_k); \\ &\Delta\mathbf{P}_{k+1}(g_1^o,\mathbf{Q}_{k+1}) = \mathbf{Q}_{k+1}(\mathbf{e}_{k+1}^q, g_1^o\varphi_{k+1}); && \Delta\mathbf{p}_{k+1}(g_2^o,\mathbf{Q}_{k+1}) = 4g_2^o\boldsymbol{\sigma}_{k+1}^q, \end{aligned} \tag{8}$$

где значение вектора МПР $\boldsymbol{\sigma}_{k+1}^q$ определяется по кватерниону $\mathbf{Q}_{k+1}$ явными соотношениями.

На завершающем этапе фильтрации дискретных измерений ИИМ с компенсацией его дрейфа выполняется формирование цифровых выходных сигналов СОУП – значений кватерниона $\hat{\mathbf{\Lambda}}_l$, которые получаются по значениям вектора МПР $\hat{\boldsymbol{\sigma}}_l$ с помощью преобразования $\hat{\boldsymbol{\sigma}}_l \Rightarrow \hat{\mathbf{\Lambda}}_l$, и вектора угловой скорости $\hat{\boldsymbol{\omega}}_l$, $l \in \mathrm{N}_0$ с заданным периодом $T_p > T_q$, кратном периоду T_q. Погрешности применяемой фильтрации выходных сигналов СОУП представлены на рис. 2 черными линиями.

Несмотря на то, что описанная дискретная фильтрация измеряемых сигналов ИИМ может выполняться по схеме параллельных вычислений, здесь возникает фиксированное временное запаздывание $T_d = 4T_q$ относительно моментов времени t_k. Авторский метод аналитического учета такого запаздывания при синтезе цифровых алгоритмов управления ориентацией КА представлен в [20].

Юстировка и калибровка масштабного коэффициента

Как отмечено выше, юстировка СОУП (вычисление оценки $\hat{\mathbf{S}}^{\Delta}$) и определение оценки $\hat{m}$ погрешности m масштабного коэффициента выполняются off-line на основе сопоставления значений вектора угловой скорости $\boldsymbol{\omega}$ спутника, которые восстанавливаются автономно по сигналам ИИМ (векторы $\hat{\boldsymbol{\omega}}^{\mathrm{g}}$) и АС (векторы $\hat{\boldsymbol{\omega}}^{\mathrm{a}}$), полученных в одинаковые моменты времени.

На фоне выполнения целевых задач спутника формируется группа трёх временных «скользящих окон» с номерами $j = 1,2,3$, с началом в моменты времени $m_j T_o$ и длительностью $n_j T_o$, где $n_j \in [12...18]$, причем каждое окно содержит интервал класса $\mathrm{S}\alpha$, где $|\boldsymbol{\omega}(t)| \le \omega_1^{\mathrm{m}}$, выполняются рекуррентная калибровка по сигналам астрономической системы и компенсация дрейфа $\mathbf{b}^{\mathrm{g}}$ ИИМ. В каждом j-ом окне длительностью $n_j T_o$: (i) явным дифференцированием полиномов $\hat{\mathbf{i}}_k^{\mathrm{g}\omega}(\tau)$ в базисе $\mathbf{G}$ получаются оценки вектора угловой скорости $\hat{\boldsymbol{\omega}}_l^{\mathrm{g}}$ с периодом T_p; (ii) по

измеренным АС значениям кватерниона $\mathbf{\Lambda}^{\mathrm{a}}_{\mathrm{m}\,k}$ формируется набор значений вектора МПР $\boldsymbol{\sigma}^{\mathrm{a}}_{\mathrm{m}k}$ и выполняется его аппроксимация по МНК полиномом 5-го порядка $\hat{\boldsymbol{\sigma}}^{\mathrm{a}}_{j}(\tau_j)$ для соответствующего локального времени $\tau_j = t - m_j T_o \in [0, n_j T_o]$, $j = 1,2,3$; (iii) по обратному уравнению для вектора МПР определяется аналитическая зависимость оценки $\hat{\boldsymbol{\omega}}^{\mathrm{a}}_{j}(\tau_j)$ вектора угловой скорости внутри j-го окна и формируются значения векторов $\hat{\boldsymbol{\omega}}^{\mathrm{a}}_{l}$ с периодом T_p. Далее с помощью переиндексации формируются два массива значений векторов $\hat{\boldsymbol{\omega}}^{\mathrm{g}}_{i}$ и $\hat{\boldsymbol{\omega}}^{\mathrm{a}}_{i}$ в моменты времени $t = t_i$ с периодом T_p. Для этих векторов вычисляются значения модулей $\hat{\omega}^{\mathrm{g}}_{i} = |\hat{\boldsymbol{\omega}}^{\mathrm{g}}_{i}|$, $\hat{\omega}^{\mathrm{a}}_{i} = |\hat{\boldsymbol{\omega}}^{\mathrm{a}}_{i}|$ и ортов $\hat{\mathbf{e}}^{\mathrm{g}}_{\omega i} = \hat{\boldsymbol{\omega}}^{\mathrm{g}}_{i} / \hat{\omega}^{\mathrm{g}}_{i}$, $\hat{\mathbf{e}}^{\mathrm{a}}_{\omega i} = \hat{\boldsymbol{\omega}}^{\mathrm{a}}_{i} / \hat{\omega}^{\mathrm{a}}_{i}$, если $\hat{\omega}^{\mathrm{g}}_{i} > 0$ и $\hat{\omega}^{\mathrm{a}}_{i} > 0$.

Задача юстировки сводится к определению взаимной ориентации ортогональных базисов $\mathbf{A}$ и $\mathbf{G}$ на основе вычисленных значений ортов $\mathbf{a}_i = \hat{\mathbf{e}}^{\mathrm{a}}_{\omega i}$ в базисе $\mathbf{A}$ и ортов $\mathbf{g}_i = \hat{\mathbf{e}}^{\mathrm{g}}_{\omega i}$ в базисе $\mathbf{G}$, что формулируется так: найти оптимальную оценку $\hat{\mathbf{S}}^{\Delta}$ ортогональной матрицы $\mathbf{S}^{\Delta}$ с определителем, равным +1, которая минимизирует критерий $L = \frac{1}{2}\Sigma\alpha_i |\mathbf{g}_i - \mathbf{S}^{\Delta}\mathbf{a}_i|^2$, где неотрицательные числа α_i являются весовыми коэффициентами с условием $\Sigma\alpha_i = 1$. Как известно, решением этой задачи является оптимальный кватернион $\mathbf{\Lambda}^{\Delta}$, который взаимно-однозначно связан с искомой ортогональной матрице $\mathbf{S}^{\Delta}$ и определяется как собственный вектор матрицы $\mathbf{K}$ с наибольшим собственным значением q_{m}: кватернион $\mathbf{\Lambda}^{\Delta}$ определяется по алгоритму

$$\mathbf{B} \equiv \Sigma\, \alpha_i \mathbf{g}_i \mathbf{a}_i^{\mathrm{t}}; \quad \mathbf{Q} \equiv \mathbf{B} + \mathbf{B}^{\mathrm{t}}; \mathbf{z} \equiv \Sigma\, \alpha_i \mathbf{g}_i \times \mathbf{a}_i; \quad \mathbf{K} = \begin{bmatrix} \mathrm{tr}\mathbf{B} & \mathbf{z}^{\mathrm{t}} \\ \mathbf{z} & \mathbf{Q} - \mathbf{I}_3 \mathrm{tr}\mathbf{B} \end{bmatrix}; \quad \mathbf{K}\mathbf{\Lambda}^{\Delta} = q_{\mathrm{m}} \mathbf{\Lambda}^{\Delta}. \tag{9}$$

Соотношения (9) представляют известный алгоритм QUEST (Quaternion Estimation) для получения оптимальной оценки $\hat{\mathbf{\Lambda}}^{\Delta}$ кватерниона $\mathbf{\Lambda}^{\Delta}$ по векторным измерениям [21].

Для калибровки погрешности m масштабного коэффициента вычисляется последовательность значений $m_i = 1 - (\hat{\omega}^{\mathrm{g}}_{i} / \hat{\omega}^{\mathrm{a}}_{i})$, искомая оценка $\hat{m}$ получается их осреднением и применяется как $\hat{m}_k$ до завершения следующей калибровки.

На временных интервалах, не принадлежащих классу $\mathrm{S}\alpha$, астрономическая коррекция выключается и оценка углового положения КА продолжает выполняться с использованием прогноза вариации оценки $\hat{\mathbf{b}}^{\mathrm{g}}$ вектора дрейфа $\mathbf{b}^{\mathrm{g}}$. Такой прогноз формируется на основе анализа тренда этой вариации на предыдущем интервале работы СОУП с астрономической коррекцией. При этом выполняется аппроксимация по МНК набора оценок вектора дрейфа ИИМ, сформированных в последнем «скользящем окне» ($j = 3$), векторным полиномом 2-го порядка, на основе которого реализуется прогноз вариации вектора дрейфа ИИМ для моментов времени $t_k > (m_3 + n_3)T_o$ вплоть до начала очередного временного интервала класса $\mathrm{S}\alpha$, когда включается астрономическая коррекция ИИМ.

Компьютерная имитация работы системы определения углового положения

Пусть для КА землеобзора на солнечно-синхронной орбите высотой 720 км имеется задание на съемку окрестностей российских столиц (рис. 3) в такой последовательности временных интервалов при отсчете времени от момента прохождения восходящего узла орбиты: ориентация КА в орбитальной системе координат $\forall t \in [0,870)$ с; ПМ1 длительностью 40 с $\forall t \in [870,910)$ с; маршрут М1 съемки Москвы с выравниванием продольной скорости движения изображения длительностью 40 с $\forall t \in [910,950)$ с, ширина полосы захвата 41 км, длина 271 км; ПМ2 длительностью 55 с $\forall t \in [950,1005)$ с; маршрут М2 трассовой съемки Санкт-Петербурга длительностью 15 с $\forall t \in [1005,1020)$, ширина полосы захвата 40 км, длина 102 км.

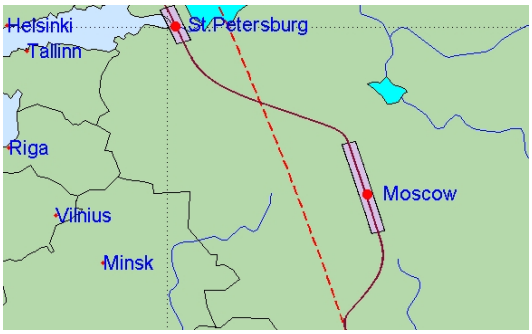


Рис. 3. Маршруты съемки двух столиц на карте

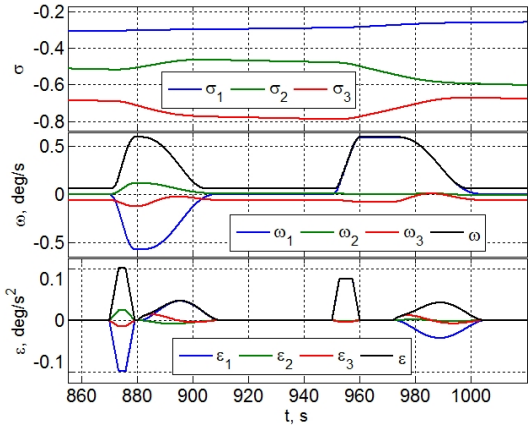


Рис. 4. Векторный сплайновый закон наведения

На рис. 4 представлен закон наведения КА в инерциальном базисе для выполнения указанного задания, где цветами выделены компоненты векторов $\boldsymbol{\sigma}(t)$, $\boldsymbol{\omega}(t)$ и $\boldsymbol{\varepsilon}(t)$ – синим цветом по крену, зеленым по рысканию и красным по тангажу, а модули векторов $\boldsymbol{\omega}(t)$ и $\boldsymbol{\varepsilon}(t)$ отмечены черным цветом. На рис. 5 приведены погрешности оценки $\delta\hat{\mathbf{b}}_k^{\mathrm{g}} = \mathbf{b}_k^{\mathrm{g}} - \hat{\mathbf{b}}_k^{\mathrm{g}}$ вектора дрейфа ИИМ, когда тестовый вектор дрейфа $\mathbf{b}^{\mathrm{g}} = \{1, -0.8, 0.3\}$ угл.сек/с восстанавливается с нулевых значений за 60 секунд.

В моменты времени t_l с периодом $T_p = 0.125$ с (частота 8 Гц) погрешность оценки ориентации представляется вектором $\boldsymbol{\delta}_l \equiv \delta\boldsymbol{\phi}_l = 4\boldsymbol{\sigma}_l^{\mathrm{e}}$. Цифровая фильтрация вектора $\boldsymbol{\delta}_l$ с частотой 8 Гц дает вектор $\boldsymbol{\delta}^{\mathrm{f}}(t)$, его компоненты изображены на рис. 6. Прогноз вариации оценки $\hat{\mathbf{b}}^{\mathrm{g}}$ без астрономической коррекции выполнялся по набору оценок $\hat{\boldsymbol{\omega}}^{\mathrm{g}}$ и $\hat{\boldsymbol{\omega}}^{\mathrm{a}}$, сформированных в последнем «скользящем окне» длительностью 20 секунд. С помощью разработанных процедур юстировки и калибровки СОУП получены оценки компонентов вектора «скосов» с точностью 3 угл.сек и $\hat{m}$ с точностью 0.025 %.

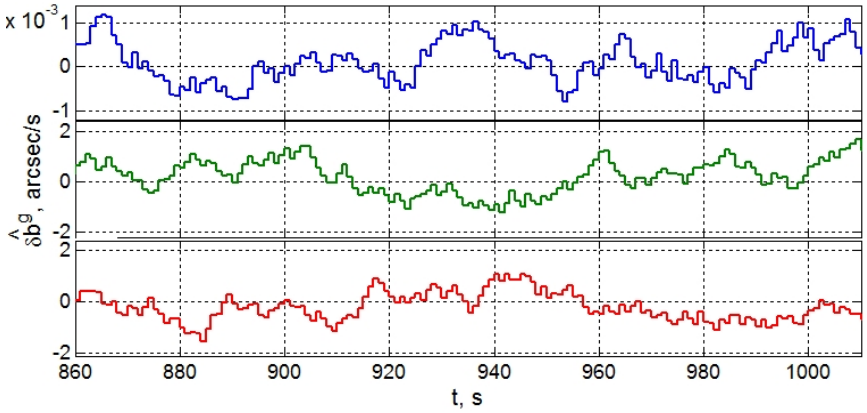


Рис. 5. Погрешности оценки вектора дрейфа ИИМ в базисе $\mathbf{G}$

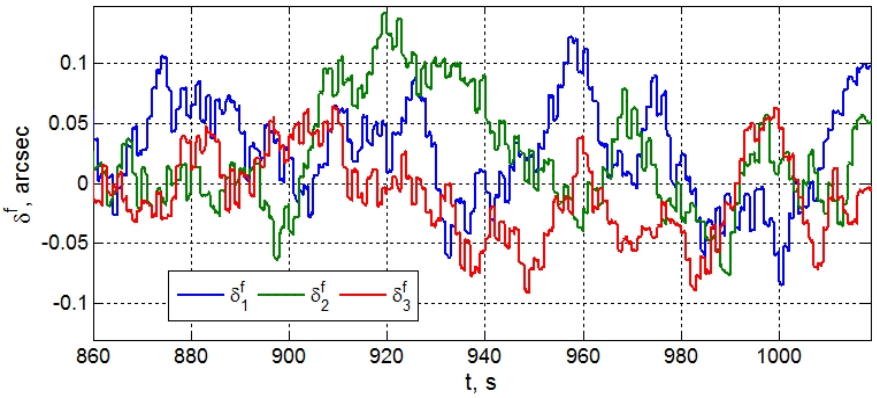


Рис. 6. Угловые погрешности СОУП при определении ориентации КА

Заключение

Кратко представлены оригинальный метод дискретной двухпроходной фильтрации измерений ИИМ, методики калибровки и юстировки астроинерцильной системы для определения кинематических параметров ориентации спутника землеобзора. Приведены результаты компьютерной имитации, демонстрирующие эффективность разработанных алгоритмов.

Работа поддержана грантами РФФИ № 17-48-630637 и № 17-08-01708.

ЛИТЕРАТУРА

1. Ишлинский А.Ю. Ориентация, гироскопы и инерциальная навигация. М.: Наука. 1976.
2. Кульба В.В., Микрин Е.А., Павлов Б.В., Платонов В.Н. Теоретические основы проектирования информационно-управляющих систем космических аппаратов. М.: Наука. 2006.
3. Бранец В.Н., Шлыглевский И.П. Введение в теорию бесплатформенных инерциальных навигационных систем. М.: Наука. 1992.
4. Titterton D.H., Weston J.L. Strapdown inertial navigation technology. 2nd ed. Cornwall: MPG Books Ltd. 2004.
5. Анучин О.Н., Комарова И.Э., Порфирьев Л.Ф. Бортовые системы навигации и ориентации искусственных спутников Земли. СПб: Концерн «ЦНИИ Электроприбор». 2004.
6. Матвеев В.В., Распопов В.Я. Основы построения бесплатформенных инерциальных навигационных систем. СПб: Концерн «ЦНИИ Электроприбор». 2009.
7. Somov Ye.I. Multiple algorithms for filtration, integration and calibration of a strap-down inertial system for a spacecraft attitude determination. Proceedings of 16th Saint Petersburg International Conference on Integrated Navigation Systems. 2009, pp. 110-112.
8. Somov Ye.I., Butyrin S.A. Digital signal processing, calibration and alignment of a strap-down inertial system for attitude determination of an agile spacecraft. Proceedings of 17th Saint Petersburg International Conference on Integrated Navigation Systems. 2010, pp. 81-83.
9. Pittelkau M.E. Kalman filtering for spacecraft system alignment calibration. Journal of Guidance, Control and Dynamics. 2001. Vol. 24, no. 6, pp. 1187-1195.
10. Bortz J.E. A new mathematical formulation for strapdown inertial navigation. IEEE Transactions on Aerospace and Electronic Systems. 1971. Vol. 7, no. 1, pp. 61-66.
11. Gusinsky V.Z., Lesyuchevsky M.V., Litmanovich Yu.A., Mussoff H., Schmidt G.T. New procedure for deriving optimized strap-down attitude algorithms. Journal of Guidance, Control and Dynamics. 1997. Vol. 20, no. 4, pp. 673-680.
12. Бабич О.А. Исследование некоммутативных поворотов в алгоритмах ориентации БИНС методом аксоидов // Труды МИЭА. 2013. Вып. 6. С. 18-34.
13. Yong K., Jo S., Bang H. A modified Rodrigues parameter-based nonlinear observer design for spacecraft gyroscope parameters estimation. Transaction of the Japan Society for Aeronautical and Space Science. 2012. Vol. 55, no. 5, pp. 313-320.
14. Somov Ye. I., Butyrin S.A., Skirmunt V.K. In-flight alignment calibration a space telescope and a star tracker cluster. Proceedings of the jubilee 15th Saint Petersburg International Conference on Integrated Navigation Systems. 2008, pp. 139-143.
15. Somov Ye.I., Butyrin S.A. In-flight alignment of a space telescope and a star tracker cluster at a scanning observation of the Earth marked objects. Proceedings of 19th Saint Petersburg International Conference on Integrated Navigation Systems. 2012, pp. 242-244.
16. Kwakernaak H., Sivan R. Linear Optimal Control Systems. Wiley-Interscience. 1972.
17. Сомов Е.И. Структурные свойства и канонические представления линейных систем управления. Иркутск: Изд-во Иркутского гос. ун-та. 1993.
18. Shampine L.F. Some practical Runge-Kutta formulas. Mathematics of Computation. 1986. Vol. 46, no. 173. pp. 135-150.
19. Mathews J.H., Fink K.D. Numerical methods using MATLAB. 3rd ed. New York, Prentice Hall. 1999.
20. Сомов Е.И. Робастная стабилизация упругих космических аппаратов при неполном дискретном измерении и запаздывании в управлении // Известия РАН. Теория и системы управления. 2001. № 2. С. 124-143.
21. Markley F.L., Mortari D. Quaternion attitude estimation using vector observations. Journal of the Astronautical Sciences. 2000. Vol. 48, pp. 359-380.

Ye.I. Somov, S.A. Butyrin, T.E. Somova (Samara State Technical University, Samara)

Algorithms for calibration and a discrete signal filtering in astroinertial system to determinate the orientation of a maneuvering land-survey satellite

Algorithms for filtering measurements, calibration and alignment of astroinertial system for determining kinematic parameters of the orientation for a maneuvering land-survey satellite are presented.

КОДИРОВАНИЕ АУДИО- И ВИДЕОСИГНАЛОВ С ПОМОЩЬЮ АДАПТИВНОГО КОДИРОВАНИЯ

С. И. Томашевич[1], **А. М. Зенкин**[2]
(*Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики,
С.-Петербург, пр. Кронверкский, 49;
Институт Проблем Машиноведения Российской Академии Наук,
С.-Петербург, Васильевский остров, Большой проспект, 61,
tomashevich.stanislav@gmail.com, zenkinartem1997@gmail.com*)

Аннотация

В работе предлагается применение адаптивного алгоритма кодирования для обработки и передачи аудио- и видеосигналов. Показано, что аудиосигнал представляет собой переходный процесс, к которому может быть применено адаптивное кодирование. Для формирования кодируемых видеосигналов использовано разбиение цвета каждого пикселя на три компоненты (красную, зеленую и синюю). Предложенный подход позволяет кодировать сигналы в реальном времени без их предварительного полного анализа. В зависимости от скорости кодирования и характера сигнала возможна экономия объема передаваемой информации в результате работы алгоритма.

Ключевые слова: адаптивное кодирование, аудиосигнал, видеосигнал, наблюдатель

Введение

В последнее время с развитием беспроводной связи появляется все больше возможностей для автономных систем, осуществляющих взаимодействие с окружающей средой, друг с другом и с пользователями. Для этого обмена информацией необходимо наладить канал связи, который может иметь низкую пропускную способность. Примерами сигналов, которые могут быть переданы между системами, могут служить навигационные данные о положении, состоянии, различная информация, получаемая с датчиков, в том числе аудио- и видеосигналы. В связи с этим становится актуальным использование такого кодирования, которое сможет в реальном времени обрабатывать сигналы для последующей передачи через канал данных с ограниченной пропускной способностью.

Предлагаемый для данной задачи адаптивный алгоритм кодирования информации относится к числу алгоритмов, работающих в реальном времени. Большинство распространенных алгоритмов используют полностью известный сигнал, тем самым имея возможность анализировать особенности кодируемого сигнала. Такой подход позволяет кодировать данные с минимальными потерями [1-7].

Предлагаемый в настоящей работе алгоритм основан на алгоритме с использованием квантования с памятью [8, 9], который, в свою очередь, основан на методе, предложенном для статического кодера [10]. Альтернативное применение схожих алгоритмов может быть найдено, например, в задачах кодирования навигационной информации летательных аппаратов [11, 12].

Таким образом, в связи со схожей природой аудио- и видеосигналов возможно применение к ним методов адаптивного кодирования. Предлагаемый доклад посвящен способам и результатам этого применения.

Алгоритм адаптивного кодирования

Приведем принцип работы алгоритма адаптивного кодирования для кодирования и декодирования информации, содержащейся в сигнале [1]. Пусть $y[k]$ – скалярный сигнал, подлежащий передаче по цифровому каналу связи в дискретные моменты времени $t_k = kT$, где T – интервал дискретности, k – номер шага. При двоичном кодировании выполняется соотношение

[1] Аспирант, научный сотрудник.
[2] Студент.

$R = T^{-1}$, где R – скорость передачи данных, выраженная в битах в секунду. Воспользуемся бинарным статическим квантователем:

$$q[k] = M[k]\sigma[k], \tag{1}$$

где $M[k]$ – уровень квантования, а $\sigma[k] = \text{sign}(y[k] - \hat{y}[k])$, где $\hat{y}[k]$ – оценка сигнала $y[k]$, которая будет сформирована в дальнейшем. Значение σ представляется однобитовым символом входного алфавита. Существенным является предположение, что кодер и декодер работают синхронно, а также, что передача данных в канале связей осуществляется без потерь. В качестве алгоритма настройки уровня квантования воспользуемся отличным от предложенного в работах [12, 13] алгоритмом:

$$\begin{aligned} &\lambda[k] = (\bar{\sigma}|k| + \bar{\sigma}|k-1| + \bar{\sigma}|k-3|)/3, \\ &M|k+1| = m_c + (|\lambda[k]| + 0.5)M|k|, \\ &\lambda[0] = 0, M[0] = M_0. \end{aligned} \tag{2}$$

Для оценивания кодируемого сигнала в кодер и декодер вводится наблюдатель следующего вида:

$$\begin{cases} \hat{y}|k+1| = \hat{y}|k| + T\hat{V}|k| + l_1\sigma|k|, \\ \hat{V}|k+1| = \hat{V}|k| + l_2\sigma|k|, \end{cases} \tag{3}$$

где $\hat{V}|k|$ – оценка скорости сигнала $y[k]$, а l_1, l_2 – параметры наблюдателя.

Параметр $\lambda[k]$ характеризует разброс изменений кодируемого сигнала, что дает возможность определить требуемое направление изменения параметра $M[k]$. Диапазон квантования увеличивается, если кодируемый сигнал изменяется быстро, и наоборот, уменьшается при малых изменениях этого сигнала. Такое поведение позволяет “оценить” новое кодируемое значение и в дальнейшем получить его на декодере. Параметр m_c определяет минимальный уровень квантования. Декодер, получая значения $\sigma[k]$, настраивает уровень квантования аналогично по алгоритму (2). Вычислив текущее значение (1), которое рассчитывается как $q[k] = y[k] - \hat{y}[k]$, декодер использует восстановленное значение $y[k]$ для оценивания передаваемого сигнала с помощью наблюдателя (3), в который закладывается модель динамики кодируемого сигнала.

Кодирование аудиосигнала

Продемонстрируем пример работы по кодированию аудиосигнала с использованием предложенного выше алгоритма адаптивного кодирования. В качестве исходного сигнала был взят MP3 аудиофайл с двухканальным кодированием, при котором каналы исходного стереосигнала кодируются независимо друг от друга. Файл содержит запись человеческой речи, при кодировании которой допустимы потери в качестве без потери возможности различить фразу. Важно отметить, что данный формат аудиофайлов уже использует алгоритм сжатия с потерями с использованием психоакустики для полного сигнала. Размер исходного файла составляет 1124120 бит с частотой дискретизации 44100 Гц. Без учёта служебной информации, которая включает в себя отступы, заголовки для каждого фрагмента данных и метки в границах MP3 аудиофайла, размер файла будет составлять 734676 бит. Для демонстрации используется моно-сигнал, то есть один из каналов, размер которого составляет $V = 367338$ бит. Применим предложенный алгоритм адаптивного кодирования к рассматриваемому аудиосигналу с частотой дискретизации равной исходному аудиосигналу. Результат фрагмента моделирования приведен на рисунке 1.

На рисунке 1 сплошной тонкой линией показана исходная аудиозапись голоса, толстой пунктирной линией – закодированный сигнал. На рисунке 2 показана ошибка кодирования $e[k] = y[k] - \hat{y}[k]$, характеризующая потери. Видно, что в моменты времени с 0.05 по 0.1 секунд потери практически совпадают с уровнем кодируемого сигнала. На данном временном интер-

вале исходный сигнал содержит высокочастотный шум, не содержащий полезной информации. Таким образом, использование данного алгоритма позволяет фильтровать высокие частоты исходного сигнала в зависимости от выбранной частоты дискретизации алгоритма кодирования. Размер полученного сигнала составляет $\widehat{V} = 256895$ бит, что составляет приблизительно 70% от размера исходного аудиосигнал.

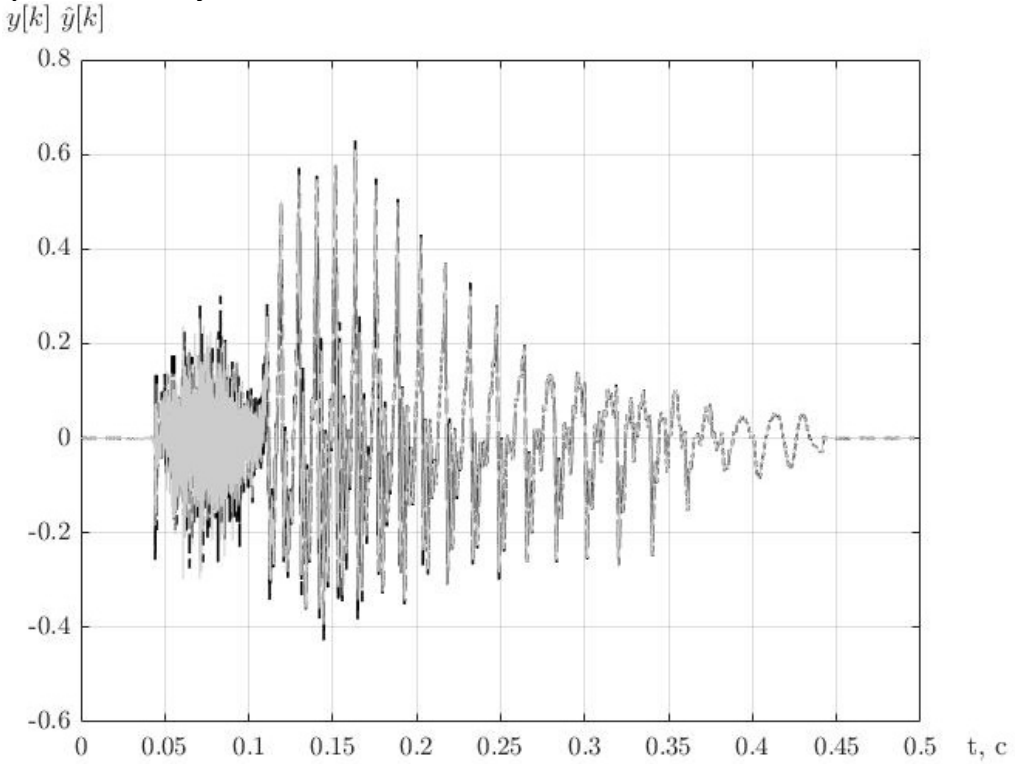


Рис. 1. Исходный и закодированный сигнал на частоте дискретизации 44100 Гц

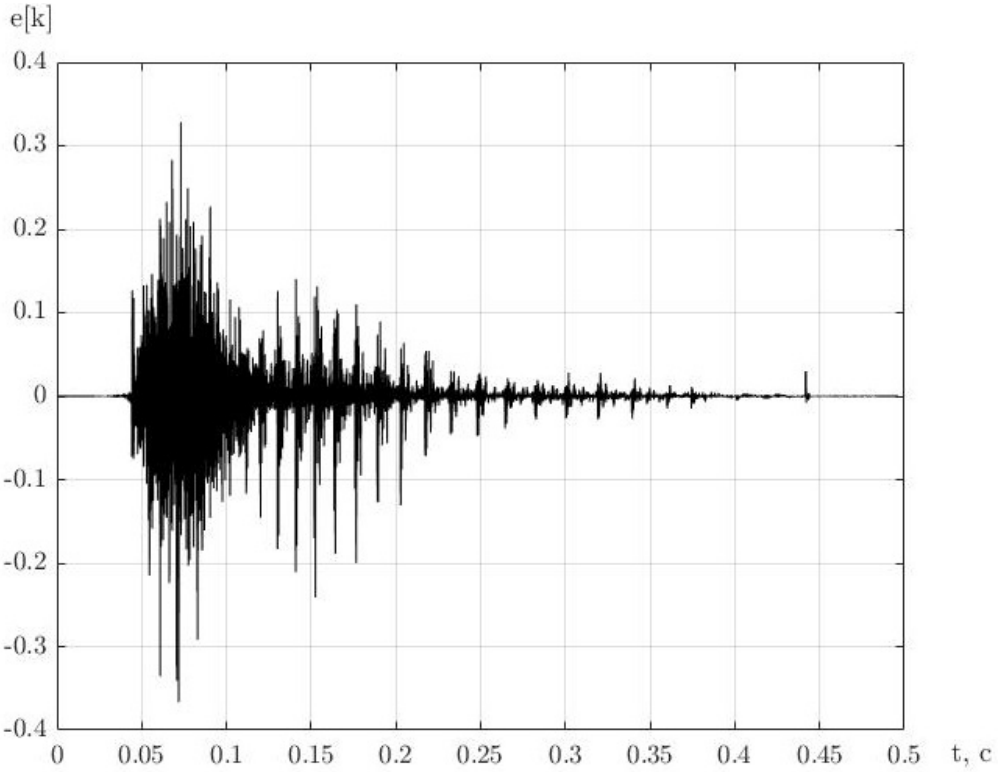


Рис. 2. Ошибка кодирования на частоте дискретизации 44100 Гц

Будем понижать частоту кодирования, пока субъективно фраза не станет неразличима на слух. В качестве критериев качества кодирования рассмотрим соотношение размеров исходного и закодированного сигналов, а также уровень потерь, который определим с помощью выражения:

$$\gamma = \sum_{k=1}^{k_f} \frac{1}{2}(e^2[k] + e^2[k-1])T, \quad (4)$$

где k_f – номер последнего измерения. Выражение (4) описывает численное интегрирование квадрата ошибки $e[k]$ с помощью метода трапеций. Результаты эксперимента приведены на рисунках 3 и 4.

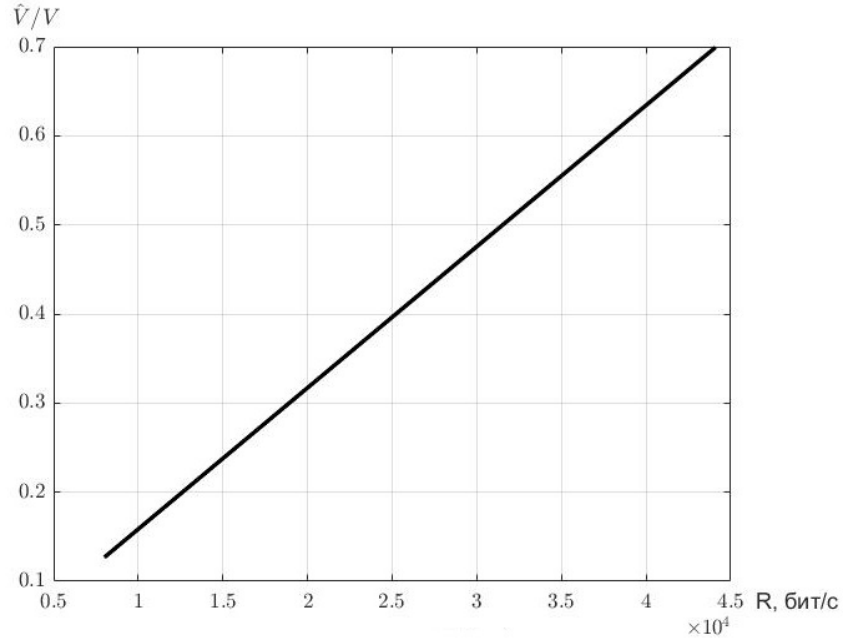


Рис. 3. Зависимость отношения закодированного аудиосигнала к исходному от скорости кодирования

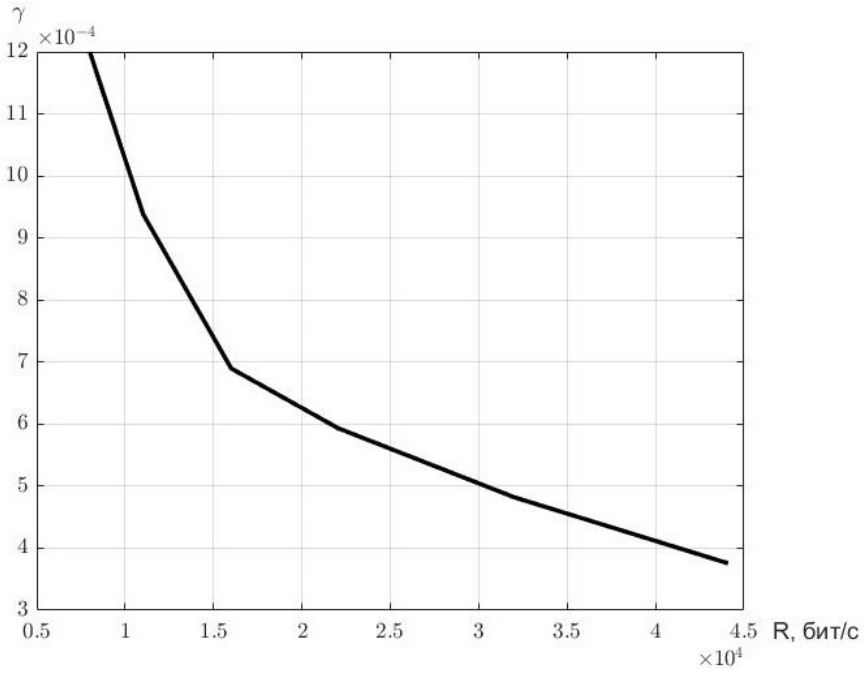


Рис. 4. Зависимость качества аудиосигнала от скорости кодирования

На рисунках 3 и 4 показано влияние частоты кодирования сигнала на отношение размера закодированного сигнала к исходному ($\widehat{V} / V$) и уровень потерь γ. Таким образом, уменьшение скорости кодирования аудиосигнала и последующей передачи через канал связи позволяет добиться уменьшения объема передаваемой информации с потерей качества, которая допустима для рассматриваемых сигналов.

Кодирование видеосигнала

Для кодирования видеосигнала предлагается разделить все изображение на пиксели, каждый из которых описывается тремя цветами – красным, зеленым и синим. Для представления каждого цвета используется один октет, значения которого обозначаются для удобства целыми числами от 0 до 255 включительно, где 0 — минимальная, а 255 — максимальная интенсивность. Последовательность кадров видеосигнала для каждого пикселя определяет процесс изменения интенсивностей каждой из компонент цвета. Применим описанный выше алгоритм к каждому цвету в пикселе для всех пикселей изображения. Очевидно, что для видеосигнала без резкой смены кадров процесс смены интенсивности трех цветов может быть медленно меняющимся, что позволит быстрее по времени производить его оценку. Таким образом, кодирование пикселя за один шаг дискретизации осуществляется тремя битами. Для кодирования полноценного изображения необходимо потратить 24 бита на пиксель. Очевидно, что размер закодированного сигнала будет всегда меньше исходного вплоть до частоты кодирования в восемь раз большей частоты кадров в секунду кодируемого видеосигнала.

Для демонстрации процесса кодирования видеосигнала, используем произвольный видеофрагмент, преобразованный в набор кадров. Результаты кодирования для компонент цвета одного из пикселей показаны на рисунке 5.

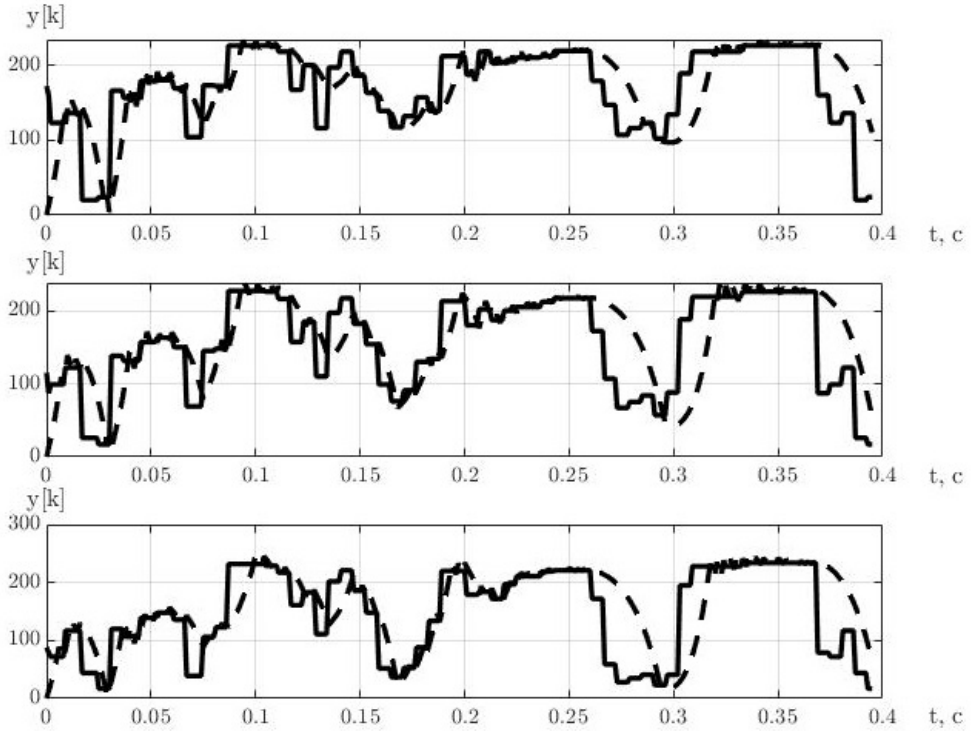


Рис. 5. Закодированные красная, зеленая и синяя составляющие на частоте дискретизации 1000 Гц

На рисунке 5 сплошной линией обозначены интенсивности компонент цвета пикселя в исходном видеосигнале, а пунктиром – в закодированном.

Для демонстрации процесса кодирования приведем набор последовательных кадров, показанных на рисунке 6.

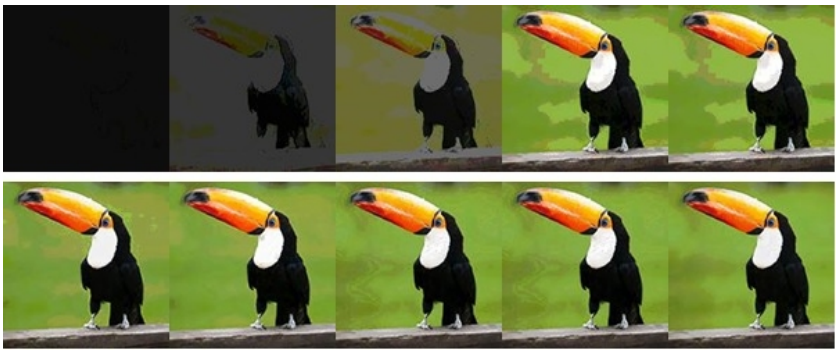


Рис. 6. Промежуточные кадры кодирования статического изображения с исходным изображением

На рис. 6 приведены девять кадров, полученных в процессе кодирования статического изображения, последний десятый кадр представляет собой кодируемое изображение. Из рисунка можно видеть, что изображение становится различимо уже на первых шагах кодирования, а в дальнейшем лишь настраивается более точно по цвету. Такая особенность позволяет сразу распознавать изображение из видеосигнала, не дожидаясь окончания переходного процесса.

Заключение

В настоящем докладе рассмотрено применение алгоритма адаптивного кодирования для аудио- и видеосигналов в условиях ограниченной пропускной способности канала связи. Предложенный метод позволяет в системах реального времени использовать низкую скорость кодирования с потерей качества, субъективно незаметной и неважной информации. Эффективность предложенного метода продемонстрирована на примере кодирования аудиосигнала, содержащего человеческую речь, произвольного видеосигнала и статического изображения. В дальнейшем есть возможность улучшить качество кодирования сигналов с помощью выбора другой модели, описывающей динамику кодируемых сигналов. Таким образом, одним из вариантов развития исследований является исследование природы динамики аудио- и видеосигналов и соответствующая модификация предложенного алгоритма.

ЛИТЕРАТУРА

1. Jayant N., Johnston J., Safranek R. Signal compression based on models of human perception // Proceedings of the IEEE. 1993. Vol. 81, N 10. P. 1385–1422.
2. Chen Z., Guillemot C. Perceptually-friendly H.264/AVC video Coding // Image Processing (ICIP), 2009 16th IEEE International Conference on. 2009. P. 3417–3420.
3. Chen Z., Guillemot C. Perceptually-Friendly H.264/AVC Video Coding Based on Foveated Just-Noticeable-Distortion Model // IEEE Transactions on Circuits and Systems for Video Technology. 2010. Vol. 20, N 6. P. 806–819.
4. Noll P. MPEG digital audio coding // IEEE Signal Processing Magazine. 1997. Vol. 14, N 5. P. 59–81.
5. Кулешов С. В. Временной анализ кодеков H.264 // Известия высших учебных заведений. Приборостроение. 2017. Vol. 60, N 11. P. 1092–1095.
6. Рубина И. С., Тропченко А. Ю. Исследование алгоритмов кодирования преобразованием в задачах сжатия кадров видеопоследовательности // Известия высших учебных заведений. Приборостроение. 2012. Vol. 55, N 10. P. 26–30.
7. Юрков К. В., Петров, С. Е. О двух подходах к параметрическому кодированию стереосигнала // Известия высших учебных заведений. Приборостроение. 2012. Vol. 55, N 1. P. 57–62.
8. Andrievsky B., Fradkov A., Peaucelle D. State estimation over the limited-band communication channel for pitch motion control of LAAS Helicopter benchmark // Proc. 17th IFAC Symp. Aut. Contr. Aerospace (ACA'2007), Toulouse, France. 2007.
9. Fradkov A., Andrievskiy B., Evans R. Hybrid quantized observer for multi-input-multi-output nonlinear systems // Dynamics and control of hybrid mechanical systems. World Scientific Series on Nonlinear Science, Series B / Ed. by G. Leonov, H. Nijmeijer, A. Pogromsky, A. Fradkov. Singapore: World Scientific, 2010. Vol. 14. P. 89–102.
10. Goodman D., Gersho A. Theory of an adaptive quantizer // IEEE Trans. Commun. 1974. Vol. COM-22, N 8. P. 1037–1045.
11. Andrievsky B., Fradkov A. Adaptive coding for position estimation in formation flight control // Proc. IFAC Workshop Adaptation and Learning in Control and Signal Processing (ALCOSP 2010). 2010. P. 72–76.
12. Andrievsky B., Fradkov A. L. Adaptive coding for maneuvering UAV tracking over the digital communication channel // 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops, ICUMT 2014, St. Petersburg, Russia. 2014. P. 236–241.
13. Андриевский Б. Р., Фрадков А. Л. Управление и наблюдение через каналы связи с ограниченной пропускной способностью // Гироскопия и навигация. 2009. 4 (67). СПб: ГНЦ РФ – ЦНИИ «Электроприбор». С. 103–114.

РАЗРАБОТКА И ПРОГРАММНАЯ РЕАЛИЗАЦИЯ ИМИТАТОРА СИГНАЛОВ

С. Е. Половцев[1], **Д. М. Клионский**[2]
(*Санкт-Петербургский государственный электротехнический университет "ЛЭТИ" им. В.И. Ульянова (Ленина)*)

Аннотация

Доклад посвящен разработке имитатора входных сигналов в гидроакустические средства подводного наблюдения, находящиеся в выбранной акватории и решающие задачу освещения подводной обстановки в районе. В ходе тактического эксперимента имитируется движение ряда морских объектов, некоторые из которых имеют гидроакустическое вооружение на борту (антенны). Антенны могут поддерживать режимы шумопеленгования, гидролокации и обнаружения гидролокационных сигналов. Для решения поставленной задачи использован фреймворк Qt 5.7, в качестве среды разработки QtCreator 4.0.2. Основным языком программирования является C++11, для реализации корректной интеграции с базой данных будет использован SQL.

Ключевые слова: имитатор сигналов, гидроакустические средства, режим шумопеленгования, режим гидролокации.

Введение

Доклад посвящен разработке имитатора входных сигналов в гидроакустические средства подводного наблюдения, находящиеся в выбранной акватории и решающие задачу освещения подводной обстановки в районе. Имитируется движение ряда морских объектов, участвующих в тактическом эпизоде, часть морских объектов имеет на борту гидроакустическое вооружение, а, следовательно, и антенны, часть морских объектов не имеет гидроакустического вооружения на борту. Антенны могут поддерживать один или более режимов работы. Рассматриваются режимы шумопеленгования, гидролокации и обнаружения гидролокационных сигналов.

Основная информация о гидрологии морской среды и параметрах излученных активных сигналов, параметрах гидроакустических средств подводного наблюдения, параметрах шумоизлучения объектов и их отражательной способности берется из баз данных, а также сторонних программных систем. Это позволяет разработать унифицированное программное обеспечение, позволяющее имитировать входные сигналы для различных средств наблюдения.

Разработанный сигнальный имитатор представляет собой сложную программную систему, реализованную в среде программирования QtCreator на языке C++ (прототип реализован в программной системе MATLAB) и QML (реализация интерфейса СОРДиУ – системы отображения, регистрации, документирования и управления), взаимодействующую с другими программными системами и базой данных, содержащей информацию об объектах в морской акватории.

Задачи имитатора

Имитатор предназначен для решения следующих задач [1]:

1) моделирование и визуализация (во временной и частотной областях) излучаемых активных гидроакустических сигналов и отраженных сигналов на излучающих антеннах (для исследования режима гидролокации – ГЛ и режима обнаружения гидролокационных станций - ОГС);

2) моделирование и визуализация (во временной и частотной областях) сигналов пассивного шумоизлучения морских объектов (для исследования режима шумопеленгования - ШП);

3) моделирование распространения гидроакустических сигналов в морской среде и визуализация характеристик акустического поля (во временной и частотной областях);

4) моделирование и визуализация помех морской акватории;

5) вычисление и визуализация (во временной и частотной областях) сигналов в точках приема с учетом среды распространения (для исследования режимов гидролокации, шумопеленгования и обнаружения гидролокационных станций - ОГС).

[1] Инженер-программист.
[2] Кандидат технических наук, доцент.

Для решения этих задач осуществляется задание частоты дискретизации для расчета частотных сеток сигнала в точках излучения и приема, а также расчета частотных характеристик акустического поля; задание размерности ДПФ (дискретного преобразования Фурье) для вычисления спектра сигналов; задание эквивалентных радиусов объектов в морской акватории для определения уровней отраженного сигнала;

Предусматривается сохранение промежуточных и окончательных результатов моделирования и расчетов для дальнейшего использования.

Сигналы в имитаторе

Сигнал формируется в точке излучения и в точке приема для каждого импульса, для полной посылки и серии посылок с учетом интервалов времени между импульсами и с учетом интервалов времени между посылками.

В составе имитатора формируются следующие типы сигналов [2]:

1) тональный сигнал
2) сигнал с гиперболической частотной модуляцией (ГЧМ-сигнал)
3) сигнал с линейной частотной модуляцией (ЛЧМ-сигнал)
4) сигнал с параболической частотной модуляцией (ПЧМ-сигнал)
5) шумоподобные сигналы

Особенности реализации режима ОГС в имитаторе

При имитации излучения сигнала из объекта рассчитывается отношение сигнал/помеха на входе приемной антенны на всех объектах из состава маневренной составляющей, рассчитанное в момент прихода сигнала на объекты. ОСП рассчитывается как отношение суммы спектральных плотностей мощности (СПМ) всех лучей на средней частоте сигнала к СПМ помехи на той же частоте.

Исходя из средней частоты излученного сигнала вычисляется диапазон тракта ОГС и формируется СПМ излученного сигнала в точке приема ОГС. Для получения параметров сигнала в точке приема в системе гидрологических расчетов (СГР) формируется запрос на расчет поля в соответствующем диапазоне частот. Если для диапазона 1 (1 кГц – 10 Гц) достаточно расчета СГР для шумопеленгования (ШП) и гидролокации (ГЛ), то для диапазонов 2 (10 – 30 кГц) и 3 (30 – 60 кГц) необходим новый расчет. Рассчитанные СПМ сигнала выдаются в ОГС с темпом в одну секунду и сопровождаются пеленгом на объект, излучивший сигнал, номером диапазона и кодом сигнала в базу данных (БД).

Сигнал в точке приема для ОГС определяется суммированием сигналов по лучам с учетом времен пролета по лучу. СПМ сигнала ОГС формируется в соответствии с длительностью излученного сигнала и суммируется в каждом временном интервале расчета (1 секунда).

Интеграция с модулем СГР

Для получения информации о гидрологии морской среды совместно с имитатором применяется внешний модуль «Система гидрологических расчетов». Обмен данными производится согласно модели «клиент-сервер», где имитатор является клиентом, модуль СГР – сервером, протокол обмена – TCP. После включения сервер (СГР) открывает прослушивающий TCP сокет и переходит в режим ожидания подключения со стороны клиента (имитатор). Инициатором обмена является клиент, который отправляет пакет client_inquiry, содержащий параметры запроса. В параметрах запроса содержится информация о координатах объекта-источника и объекта-приемника, количестве и варианте сортировки лучей, начальном угле, шаге по углу, числе секторов в вертикальной плоскости, частотах для расчета. После получения данного пакета сервер производит необходимые вычисления и передает результат клиенту пакетом server_answer, после чего опять переходит в режим ожидания запроса от клиента. В пакете server_answer передается информация о дистанции между объектами, числе использованных и общем количестве

лучей, соединяющих излучатель и цель, углах прихода и выхода лучей, времени полета лучей, частотах, фазах лучей в точках приема, коэффициентах передачи среды по мощности без учета сферического расхождения лучей.

Для постановки задачи эксперимента, отображения режимов ШП, ГЛ и ОГС предусмотрен интерфейс СОРДиУ, реализованный на QML. Общий вид СОРДиУ подразумевает разделение на две основные зоны: организационно-технические условия (ОТУ) и зону графической информации. В зоне ОТУ оператору имеет возможность взаимодействовать с тремя участниками тактического эксперимента: «Излучатель», «Объект» и «Приемник».

В зоне «Излучатель» оператору предоставляется возможность выбора номера излучающего объекта-участника тактического эксперимента (при этом класс, подкласс и имя объекта заполняются автоматически, согласно выбранному номеру объекта), антенны, режима и подрежима, поддерживаемых выбранной антенной. По выбранным данным оператор имеет возможность посмотреть широту, долготу, глубину, курс и скорость выбранного объекта, а также код и время излучения сигнала, угол наклона в вертикальной плоскости и частоту используемого сигнала.

В зоне «Объект» оператору предоставляется возможность выбора номера объекта-участника тактического эксперимента (может заполняться автоматически при заполненном излучателе или приемнике, при этом класс, подкласс и имя объекта заполняются автоматически, согласно выбранному номеру объекта). По выбранным данным оператор имеет возможность посмотреть широту, долготу, глубину, курс и скорость и эквивалентный радиус выбранного объекта.

В зоне «Приемник» оператору предоставляется возможность выбора номера объекта из числа приемников для выбранного излучателя (может заполняться автоматически в соответствии выбранному излучателю). Поля «Класс», «Подкласс» и «Имя» заполняются по тому же принципу, что у «Излучателя».антенны, режима и подрежима, поддерживаемых выбранной антенной. По выбранным данным оператор имеет возможность посмотреть широту, долготу, глубину, курс и скорость выбранного объекта, а также код и время излучения сигнала, угол наклона в вертикальной плоскости и частоту используемого сигнала.Оператор также имеет возможность выбрать антенну, для просмотра сигнала на графиках и установить режим и подрежим, доступные для выбранной антенны.

Зона графической информации разделяется на два столбца «Приемник» и «Излучатель» и меняет вид в зависимости от выбранного режима. При выбранном режиме ШП отображаются графики «Моделируемый сигнал», «ЧВС и амплитудный спектр» (ЧВС – частотно-временной сигнал). При выбранных режимах ГЛ и ОГС отображаются следующие графики: «Временной процесс», «Временной процесс сигнала ГЛС», «ЧВС и амплитудный спектр».

Также при загрузке программы предварительно возникает окно для установки параметров программы для дальнейшей работы. Оператор может установить количество лучей, сортировку лучей по секторам, число секторов, начальный угол, шаг по углу, конечный угол. Также оператор указывает в начале эксперимента следует ли использовать датчик случайный чисел.

Взаимодействие и связи между имитатором и некоторыми сторонними системами и модулями может быть описано на основе диаграммы компонентов.

Программная реализация имитатора может быть описана на основе диаграммы классов, показывающей набор классов и связи между ними, а также их кооперации (взаимодействия).

1) Класс Imitator предоставляет имитационную модель тактического эксперимента (данная модель включает в себя характеристики объектов-носителей и характеристики излучаемых активных и шумоподобных сигналов), управляет обновлением объектов тактического эксперимента, моделирует распространение прямых звуковых излучений (непосредственно от передающей антенны на объекте-носителе), моделирует распространение отражённых звуковых излучений и управляет формированием сигнальных данных при пересечении объекта и фронта звукового излучения (расчет сигнала на входе приемной антенны).

2) Класс Transceiver предоставляет приемопередатчик для обмена данными с диспетчером и внешними системами. По команде диспетчера инициализирует параметры тактического эксперимента для имитатора. Задает рабочий шаг имитационной модели по команде имитатора. При завершении тактического эксперимента инициализирует завершение имитационной модели.

3) Класс Sigimsord используется для связи класса Imitatori графическим интерфейсом пользователя, написанном на QML. Поддерживает разрешение контекста QML, осуществляет инициализацию и обновление данных, передаваемых с имитатора на графический интерфейс.

4) Класс Object предоставляет сущность морского объекта, моделирует шумоизлучение объекта, зависит от параметров шумоизлучения и скорости объекта, определяет настройки антенн, расположенных на борту.

5) Класс Antenna предоставляет сущность антенны гидроакустического вооружения, определяет поддерживаемые режимы и формирует приемные сигнальные данные. Агрегирует экземпляр класса RadiationFactory, для задания конфигурация излучения антенны.

6) Класс Radiation предоставляет сущность излучения активного сигнала, определяет время и позицию излучения.

7) Класс SGRTransceiver предоставляет приемопередатчик для обмена данными с диспетчером исистемой гидрологических отсчетов. Передает и принимает данные для расчета акустического поля между двумя любыми объектами тактического эксперимента.

8)КлассField предоставляет сущность акустического поля, преобразует данные, полученные от модуля СГР через SGRTransceiver, и преобразует их в структуры, удобные для обработки в имитаторе.

Заключение

Представленный имитатор позволяет осуществлять моделирование и расчет сигналов в точках излучения и приема и может использоваться при моделировании работы гидроакустических средств освещения подводной обстановки в заданной акватории и при их разработке.

ЛИТЕРАТУРА

1. Е. Л. Шейнман, А.Н. Афанасьев, М.С. Куприянов, Д.М. Клионский, С.Н. Ежов, П.Н. Максимов, А.М. Голубков Имитатор модели сигналов гидроакустических станций интегрированной системы подводного наблюдения // Известия СПбГЭТУ «ЛЭТИ», 2016. №10, С. 10-23.
2. Малышкин Г. С. Оптимальные и адаптивные методы обработки гидроакустических сигналов. Т. 1. Оптимальные методы / Г. С. Малышкин. СПб.: ОАО «Концерн «ЦНИИ «Электроприбор», 2011. 400 с.

S.E. Polovtsev, D.M. Kliosnkiy (Saint Petersburg Electrotechnical University “LETI”, Saint Petersburg)

DEvelopment and software implementation of signal simulator

The report is devoted to the development of a simulator of input signals for hydroacoustic tools of underwater observation. They are located in the fixed water area. During the tactic experiment we simulate the movement of a series of marine objects and some of them have hydroacoustic armament onboard (antennas). The antennas can support different modes including the mode of hydroacoustic signal detection. We have used the framework Qt 5.7 and QtCreator 4.0.2. The main programming language is C++11 and integration with the database is based on SQL.

Секция 6
БЕЗОПАСНОСТЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В УПРАВЛЕНИИ

ПОВЫШЕНИЕ ДОСТОВЕРНОСТИ ОЦЕНКИ СМЫСЛОВОГО НАПОЛНЕНИЯ ИНФОРМАЦИОННЫХ ОБЪЕКТОВ НА ОСНОВЕ ОБРАБОТКИ НЕПОЛНЫХ, ПРОТИВОРЕЧИВЫХ И НЕЧЕТКИХ ЗНАНИЙ

И. Б. Паращук[1], **С. А. Агеев**[2]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), 199178, Россия, С.-Петербург, 14-я линия В.О., 39,
(812) 328-71-81, parashchuk@comsec.spb.ru*)

Аннотация

В докладе рассмотрены различные виды неопределенности при реализации процессов обнаружения и противодействия нежелательной, сомнительной и вредоносной информации. Предложен подход к выработке требований по устранению неполноты и противоречивости оценки и категоризации смыслового наполнения информационных объектов. Данный подход позволяет повысить достоверность оценки и категоризации смыслового наполнения информационных объектов за счет применения методов и алгоритмов обработки неполных, противоречивых и нечетких знаний.

Ключевые слова: достоверность, неопределенность; оценка; информационный объект; знания

Введение

Сегодня перед государством и обществом стоит непростая задача поддержания эффективного комплекса мер по профилактике, обнаружению и противодействию нежелательной, сомнительной и вредоносной информации, содержащейся и распространяемой с применением сети Интернет или информационно-коммуникационных технологий. Ключевую роль при решении этой задачи играет предварительная или динамическая оценка смыслового наполнения информационных объектов. Оперативная и достоверная оценка позволяет поставить надежный заслон нежелательной и вредоносной информации. Нежелательное и сомнительное смысловое наполнение иногда называют информацией, несущей в себе контентные риски – это различные информационные ресурсы (тексты, картинки, аудио, видеофайлы, ссылки на сторонние ресурсы), содержащие противозаконную, неэтичную и вредоносную информацию.

Практический опыт в области получения достоверных оценок в рамках контент-анализа, оценок номенклатуры и содержания цифрового сетевого контента [1-9] показывает, что объективный анализ смыслового наполнения информационных объектов невозможен без обработки данных и, соответственно, получаемых на ее основе выводов и знаний в условиях неопределенности, без привлечения алгоритмов анализа противоречивых и динамичных (изменяющихся) знаний. Применение подобных методов и алгоритмов позволяет повысить достоверность оценивания и категоризации смыслового наполнения [8], а значит, в конечном итоге, повысить обоснованность принимаемых решений по обнаружению и противодействию нежелательной, сомнительной и вредоносной информации.

Этим задачам служат методы и алгоритмы обработки неполных, противоречивых и нечетких знаний. Задача объединения и совместного использования интеллектуальных моделей, алгоритмов и методов обработки неполных, противоречивых и нечетких знаний в интересах устранения неполноты и противоречивости оценки и категоризации смыслового наполнения информационных объектов и, в конечном итоге, в интересах защиты от нежелательной и вредоносной информации, распространяющейся в сети Интернет и социальных сетях, решается в такой постановке впервые. Вместе с тем, есть большой ряд работ, посвященных оцениванию и категоризации параметров различных сложных информационно-технических объектов, производственных и телекоммуникационных процессов, а также процессов принятия решения в услови-

[1] Доктор технических наук, профессор, ведущий научный сотрудник.
[2] Кандидат технических наук, научный сотрудник.

ях различного вида неопределенности. При этом под неопределенностью традиционно понимают недостаточную осведомленность лиц и систем, принимающих управленческие решения. Это неполное или неточное представление о значениях различных параметров смыслового наполнения информационных объектов в будущем, порождаемых различными причинами и, прежде всего, неполнотой или неточностью информации об условиях реализации оценки и категоризации, в том числе связанных с ними затратах и результатах.

Виды неопределенности и подходы к повышению достоверности оценки смыслового наполнения информационных объектов

Неопределенность обусловлена как недостатком знаний, необходимых для получения количественного описания протекающих в системе оценки и категоризации процессов, так и сложностью самого информационного объекта, сложностью процессов обнаружения и противодействия нежелательной, сомнительной и вредоносной информации [9]. Задача выработки требований по устранению неполноты и противоречивости оценки и категоризации смыслового наполнения информационных объектов приобретает все большую актуальность. Выработка требований по устранению неполноты и противоречивости оценки и категоризации смыслового наполнения информационных объектов происходит при существующей неопределенности внутренней и внешней среды. Внутренняя среда включает факторы, обусловленные функционированием самой интеллектуальные системы аналитической обработки цифрового сетевого контента и ее внутренними контактами. Внешняя среда представлена факторами, которые не связаны непосредственно с функционированием самой интеллектуальные системы аналитической обработки цифрового сетевого контента и имеют более широкий технический, поведенческий, стратегический и иной характер [1, 18, 19].

Предлагаемый подход к выработке требований по устранению неполноты и противоречивости учитывает условия оценки и категоризации в ходе функционирования сложной интеллектуальной системы аналитической обработки цифрового сетевого контента, условия обладания аналитиками контента, должностными лицами и сетевыми администраторами информацией, необходимой для реализации процедуры оценки и категоризации. Эти условия бывают детерминированными, вероятностными и неопределенными.

Причины неопределенности при оценке и категоризации могут быть не связаны со стохастичностью поведения объекта анализа – смыслового наполнения информационных объектов. Они могут быть связаны с опытом должностных лиц органов контроля сетевого контента, с нестабильностью условий функционирования интеллектуальных систем аналитической обработки цифрового сетевого контента, с нестабильностью параметров самого процесса оценки и категоризации. При этом обработку неопределенной информации различного типа обычно реализуют с помощью: методов теории нечетких множеств [10, 13, 18, 20, 22, 26]; методов искусственных нейронных сетей [11, 12, 14-17, 19, 21] и методов нейро-нечетких сетей [23-25].

Современные научные направления, ориентированные на вопросы безопасности информационных технологий в управлении, новые, перспективные технологии, методы и способы аналитической обработки цифрового сетевого контента сосредоточены на нескольких ключевых подходах. Это предварительная обработка сетевых информационных объектов на основе применения распределенных интеллектуальных сканеров; многоуровневый и многомодульный анализ информационных объектов на основе использования параллельных вычислений; алгоритмы адаптации и переобучения системы анализа информационных объектов; методы реализации визуальных интерфейсов для выявления и противодействия вредоносной информации, а также алгоритмы противодействия нежелательной, сомнительной и вредоносной информации.

Особое место принадлежит интеллектуальным системам аналитической обработки цифрового сетевого контента, где на ведущее место претендуют алгоритмы достоверной, оперативной и многоаспектной оценки и категоризации смыслового наполнения информационных объектов.

Именно интеллектуальные системы аналитической обработки цифрового сетевого контента смогут обрабатывать неполную, противоречивую или нечеткую информацию (знания). Вместе с тем, такие оценка и категоризация с целью профилактики, обнаружения и противодействия

нежелательной, сомнительной и вредоносной информации, осуществляются в условиях неопределенности различного характера. Например, в условиях неполноты, противоречивости или нечеткости информации (знаний) о реальном, истинном смысловом наполнении информационных объектов. Иными словами, интеллектуальные системы аналитической обработки цифрового сетевого контента реализуют алгоритмы повышения достоверности оценки смыслового наполнения информационных объектов на основе обработки неполных, противоречивых и нечетких знаний. Для этого на основе мнений экспертов и аналитических преобразований нечеткой и неполной (противоречивой) информации о среде (природная неопределенность), стратегиях противодействия (поведенческая неопределенность), целях функционирования (целевая неопределенность) и о предпочтениях, оказываемых тем или иным аспектам, которые (с точки зрения лица, принимающего решения) необходимо оценить в смысловом наполнении, осуществляется дополнительная обработка информации (знаний).

Первым реализуется этап, ориентированный на наличие в смысловом наполнении неопределенности, имеющей характер неоднозначности (нечеткости). Такие задачи традиционно решаются на основе математики нечетких множеств.

Механизм преобразований подробно описан в работе [26]. Логика и физический смысл данного этапа заключается в отказе от применения классических математических методов, ориентированных на стандарты точности и строгости, отказе от языка математики уравнений, а в использовании математического языка, основанного на понятиях нечетких множеств и лингвистических переменных. Здесь на основе мнений, предпочтений экспертов и лиц, принимающих решения, а также на основе аналитических преобразований нечеткой информации реализуются следующие подэтапы повышения достоверности: формулировка ключевых элементов процесса повышения достоверности – функций принадлежности нечетких множеств, характеризующих предварительные, неоднозначно (нечетко) заданные оценки смыслового наполнения информационных объектов; построение матриц функций принадлежности этих нечетких множеств; формирование нечеткого отношения пересечения; построение булевой матрицы отношения нечетких множеств; определение классов подобия – объема и номенклатуры смыслового наполнения информационных объектов; определение порядковой функции отношения нечетких множеств; поиск непересекающихся подмножеств локальных оценок смыслового наполнения информационных объектов и формулировка результатов нечеткого оценивания – классы подобия оценок смыслового наполнения информационных объектов и полученная система непересекающихся подмножеств таких оценок и категорий.

Вторым реализуется этап, ориентированный на наличие в смысловом наполнении неопределенности, имеющей характер недостаточности (неполноты). Такие задачи принято решать на основе математики искусственных экстраполирующих нейронных сетей, причем состояние (1 или 0) нейронов входного (промежуточного, выходного) слоя нейронной сети имеет физический смысл наличия или отсутствия в конкретном смысловом наполнении информационного объекта признаков нежелательной, сомнительной и вредоносной информации. Математический механизм преобразований и физический смысл процесса работы с недостаточной (неполной) информацией для примера однослойной нейронной сети подробно описан в работе [19].

Механизм преобразований основан на возможности обработки неполной и противоречивой экспертной информации с помощью методов теории искусственный нейронных сетей. В нашем конкретном случае, когда характер информации о наличии или отсутствии в смысловом наполнении информационного объекта признаков нежелательной, сомнительной и вредоносной информации не связан с лингвистической и физической неопределенностью, одним из действенных и математически корректных подходов к обработке неполной и противоречивой экспертной информации является использование нейросетевых алгоритмов идентификации данных. При этом, в интеллектуальных системах аналитической обработки цифрового сетевого контента, по-прежнему с учетом (но уже недостаточных, неполных) мнений, предпочтений экспертов и лиц, принимающих решения, на основе нейросетевых методов повышения достоверности реализуются следующие подэтапы: формулировка данных для нейронной сети, в частности значений начальных состояний (синаптических весов) входного слоя сети, характеризующих предварительные, недостаточно полно (неточно) заданные данные (знания) о смысловом наполнении конкретного информационного объекта; активизация входного слоя нейронной

сети; начальная инициализация нейронов выходного слоя нейронной сети; приведение нейронов входного слоя к состоянию нейронов выходного слоя; вычисление новых состояний нейронов выходного слоя нейронной сети; отработка нейронной сети до состояния стабилизации (собственно процесс нейросетевых вычислений); анализ полученных состояний нейронов выходного слоя и определение наличия или отсутствия в смысловом наполнении конкретного информационного объекта признаков нежелательной, сомнительной и вредоносной информации; формулировка результатов работы нейросетевого алгоритма повышения достоверности – полученная оценка, принадлежащая или не принадлежащая пространству признаков нежелательной, сомнительной и вредоносной информации.

В результате последовательной реализации этапов повышения достоверности оценки смыслового наполнения информационных объектов с учетом неопределенности, опираясь на функции принадлежности нечетких множеств и синаптические веса нейронных сетей, сформулированные с учетом предпочтений экспертов и текущих потребностей в количественных оценочных значениях смыслового наполнения, может быть получен ряд оценок, достоверно характеризующих, по мнению экспертов, однозначное наличие или отсутствие в конкретном информационном объекте признаков нежелательной, сомнительной и вредоносной информации.

Так, в результате последовательной реализации (в рамках интеллектуальных систем аналитической обработки цифрового сетевого контента) этапов повышения достоверности оценивания с учетом неопределенности, могут быть получены варианты непересекающихся множеств оценок смыслового наполнения, безусловно характеризующих принадлежность или не принадлежность признаков конкретного информационного объекта пространству признаков нежелательной, сомнительной и вредоносной информации.

Заключение

Таким образом, объективно существующая неопределенность условий оценки и категоризации смыслового наполнения информационных объектов, вызванная нестационарностью функционирования интеллектуальных систем аналитической обработки цифрового сетевого контента, воздействиями дестабилизирующих (часто антагонистических) факторов, внешней среды, нечеткостью целей и несогласованностью задач оценки и категоризации и др., обуславливает необходимость решения должностными лицами органов контроля сетевого контента (аналитиками, социологами, сетевыми администраторами и т.п.) задач оптимальной, достоверной оценки и категоризации смыслового наполнения информационных объектов с учетом условий не только многокритериальности, но и неопределенности.

Сущность методологии повышения достоверности оценки смыслового наполнения информационных объектов на основе обработки неполных, противоречивых и нечетких знаний опирается на общие методы, подходы и меры по профилактике, обнаружению и противодействию нежелательной, сомнительной и вредоносной информации, на анализе информационного объекта, его нежелательного и сомнительного смыслового наполнения. Проведенный анализ показывает, что на современном этапе недостаточно развиты алгоритмы оценки и категоризации в условиях нестохастической неопределенности, а именно, неоднозначности (нечеткости) и недостаточности (неполноты и противоречивости) исходной информации для объективного контент-анализа, для достоверного оценивания смыслового наполнения информационных объектов. Предложен подход на основе интеллектуальных систем аналитической обработки цифрового сетевого контента, он основан на предварительной выработке требований и на реализации алгоритмов устранения неполноты и противоречивости оценки и категоризации смыслового наполнения информационных объектов. Данный подход позволяет повысить достоверность оценки и категоризации смыслового наполнения информационных объектов за счет применения методов и алгоритмов обработки неполных, противоречивых и нечетких знаний.

Возможным направлением предполагаемых дальнейших исследований может быть синтез общей архитектуры интеллектуальной системы аналитической обработки цифрового сетевого контента в интересах защиты от нежелательной, сомнительной и вредоносной информации.

Работа выполнена при финансовой поддержке гранта РНФ (проект № 18-11-00302) в СПИИРАН.

ЛИТЕРАТУРА

1. Котенко И.В., Юсупов Р.М. Информационные технологии для борьбы с терроризмом // Защита информации. INSIDE № 2, 2009. С. 74-79.
2. Комашинский Д.В., Котенко И.В., Чечулин А.А. Категорирование веб-сайтов для блокирования веб страниц с неприемлемым содержимым // Системы высокой доступности, № 2, 2011. С.102-106.
3. Kotenko I.V., Chechulin A.A., Shorov A.V., Komashinsky D.V. Analysis and Evaluation of web Pages Classification Techniques for Inappropriate Content Blocking // P. Perner (Ed.): *14th Industrial Conference on Data Mining (ICDM 2014), Lecture Notes in Artificial Intelligence* (LNAI 8557), P. 39–54.
4. Elo S., Kyngas H. The qualitative content analysis process. // *Journal of Advanced Nursing* 62(1), 2008. P. 107-115.
5. Krippendorff K. Content Analysis: An Introduction to Its Methodology. *Thousand Oaks,* CA: Sage. 2004. 413 p.
6. Vaismoradi M., Turunen H., Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study.// *Nursing & Health Sciences*. 15 (3): 2013. P. 398-405.
7. Heydarian Pashakhanlou A. Fully integrated content analysis in international relations. // *International Relations.* 2017. 31 (4): P. 447-465.
8. Timmermans S., Iddo T. Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis. // *Sociological Theory* (30(3) ed.): 2012. P. 167-186.
9. Graneheim U. H., Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *// Nurse Education Today*. 2004. 24 (2): P. 105-112.
10. Kotenko I.B., Saenko I.B., Ageev S.A., Kopchak Y.M. Abnormal Traffic Detection in networks of the Internet of things based on fuzzy logical inference. // *XVIII International Conference on Soft Computing and Measurements* (SCM'2015) IEEE Xplore. 2015. P. 5-8.
11. Rojas R. Neural Networks. Berlin: *Springer-Verlag*. 1996. 453 p.
12. Parashchuk I.B. System Formation Algorithm of Communication Network Quality Factors using Artificial Neural Networks // 1st *IEEE International Conference on Circuits and System for Communications* (ICCSC'02). Proceedings. (St. Petersburg: SPbGTU). 2002. P. 263-266.
13. Kosko B. Neural Networks and Fuzzy Systems. A Dynamical Systems Approach to Machine Intelligence. *Englewood Cliffs: Prentice-Hall*, 1992. 346 p.
14. Kriesel D. A Brief Introduction to Neural Networks. *Cambridge: Cambridge Press*, 2010. 226 p.
15. Rojas R. Neural Networks. *Berlin: Springer-Verlag*, 1996. 453 p.
16. Muller B., Reinhardt J., Strickland M.T. Neural networks: an introduction. Number v. 1 in Physics of neural networks. *Springer,* 1995. P. 214-224.
17. Parlos A.G., Menon S.K., Atiya A.F. An algorithmic approach to adaptive state filtering using recurrent neural networks. // *IEEE Trans. Neural Networks*. 2001. Vol.12. No 6. P. 1411-1432.
18. Шапиро Д.И. Принятие решений в системах организационного управления: Использование расплывчатых категорий. М.: Энергоатомиздат, 1983, 184 с.
19. Паращук И.Б., Иванов Ю.Н., Романенко П.Г. Нейросетевые методы в задачах моделирования и анализа эффективности функционирования сетей связи. / Учебно-методическое пособие. СПб.: ВАС, 2010. 103 с.
20. Асаул А.Н., Карпов Б.М., Перевязкин М.К. и др. Модернизация экономики на основе технологических инноваций. СПб.: ИПЭВ, 2008. 606 с.
21. Баюк О.А., Браилов А.В., Денежкина И.Е. и др. Принятие финансовых решений в условиях сравнительной неопределенности. М.: Вузовский учебник. 2014. 106 с.
22. Военная системотехника и системный анализ. Модели и методы подготовки и принятия решений в сложных организационно-технических комплексах в условиях неопределенности и многокритериальности. / Под ред. Б.В. Соколова. СПб.: ВИКУ им. А.Ф. Можайского, 1999. 496 с.
23. Булдакова Т.И., Миков Д.А. Оценка информационных рисков в автоматизированных системах с помощью нейро-нечёткой модели. // Наука и образование. № 11, 2013. // [Электронный ресурс]: http://technomag.edu.ru/doc/645489.html.
24. Паращук И.Б., Михайличенко Н.В. Нейро-нечеткие модели в интересах управления данными в ЦОД и интеллектуального анализа информации в региональных телекоммуникационных сетях. // Юбилейная XV-ая Санкт-Петербургская Международная Конференция «Региональная информатика-2016 (РИ-2016)». Материалы конференции. СПб., СПОИСУ, 2016. 599 с., С. 117-118.
25. Паращук И.Б., Башкирцев А.С., Михайличенко Н.В. Анализ уровней и видов неопределенности, влияющей на принятие решений по управлению информационными системами.// Информация и космос. № 1, 2017. С. 112-120.
26. Паращук И.Б., Бобрик И.П. Нечеткие множества в задачах анализа сетей связи. СПб.: ВУС, 2001. 80 с.

I.B. Parashchuk, S.A. Ageev (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), St. Petersburg)

Increase of reliability of an estimation of semantic filling of information objects on the basis of processing incomplete, contradictory and fuzzy knowledge

Abstract. The report considers various types of uncertainty in the implementation of detection and counteraction processes of undesirable, questionable and harmful information. An approach to the development of requirements for eliminating the incompleteness and inconsistency of the assessment and categorization of the semantic content of information objects is proposed. This approach makes it possible to increase the reliability of evaluation and categorization of the semantic content of information objects through the application of methods and algorithms for processing incomplete, contradictory and fuzzy knowledge.

ОБЩАЯ АРХИТЕКТУРА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ АНАЛИТИЧЕСКОЙ ОБРАБОТКИ ЦИФРОВОГО СЕТЕВОГО КОНТЕНТА В ИНТЕРЕСАХ ЗАЩИТЫ ОТ НЕЖЕЛАТЕЛЬНОЙ ИНФОРМАЦИИ

И. В. Котенко[1], **И. Б. Паращук**[2]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН),
199178, Россия, С.-Петербург, 14-я линия В.О., д. 39, (812) 328-71-81,
parashchuk@comsec.spb.ru*)

Аннотация

В докладе рассмотрены элементы и уровни общей архитектуры интеллектуальной системы аналитической обработки цифрового сетевого контента, их функции и особенности. Эта архитектура представляет собой комплекс средств, методов, методик, моделей, алгоритмов и программных прототипов интеллектуальной аналитической обработки цифрового сетевого контента в интересах эффективного обнаружения и противодействия нежелательной, сомнительной и вредоносной информации, распространяемой в сети Интернет и в социальных сетях. Компоненты общей архитектуры найдут применение в системах родительского контроля, системах выявления и противодействия террористической и экстремистской активности в Интернете и социальных сетях, а также в системах анализа социальных сетей на выявление групповой вредоносной и нежелательной информационной активности.

Ключевые слова: архитектура; система; контент; информация; безопасность.

Введение

Стремительное внедрение глобальной сети Интернет и построенных на ее основе социальных сетей в государственную, производственно-экономическую и социально-культурную сферы современного общества является мощным стимулом их дальнейшего развития. В то же время, Интернет и социальные сети становятся одной их важнейших угроз персональной, общественной и государственной информационной безопасности.

В России и ведущих странах мира проблема борьбы с нежелательной, опасной информацией находится под пристальным государственным вниманием [1-6]. Наличие у нас в стране, а также в ведущих странах мира регуляторов, отвечающих за блокировку незаконного и нежелательного контента в Интернете и социальных сетях, не говорит о том, что проблема защиты от вредоносной и нежелательной информации решена.

Обнаружение вредоносных сайтов и сообщений и формирование «черных списков» в настоящий момент осуществляется, как правило, в ручном режиме [7-11]. Однако в настоящее время эти средства обладают недостаточно высокими показателями оперативности, полноты и точности [12]. Кроме того, эти алгоритмы плохо работают на наборах данных малого объема, в результате чего из-под контроля выпадают многие типы социальных сетей (Твиттер) и мобильные мессенджеры (Viber, WhatsApp). Наконец, анализ Интернет-контента зачастую связан с обработкой не только колоссальных объемов разнородных, но также противоречивых и изменчивых данных. В результате возникает необходимость применения методов обработки неполных, противоречивых и нечетких знаний.

В настоящее время в рамках технологии обработки больших данных известно множество различных подходов, инструментов и методов обработки структурированных и неструктурированных данных, в том числе в области защиты информации [13-15].

Решение этой проблемы связано с разработкой новых моделей, методов и методик, относящихся к новым фундаментальным, передовым и быстро развивающимся областям научного знания, таким как интеллектуальный анализ данных, обработка больших данных, обработка неполных, противоречивых и нечетких знаний и др.

Предполагается, что предлагаемая совокупность моделей, методов, методик и программных приложений позволит обосновать и разработать новую информационную технологию и, в ко-

[1] Доктор технических наук, профессор, заведующий лабораторией проблем компьютерной безопасности.
[2] Доктор технических наук, профессор, ведущий научный сотрудник.

нечном итоге, интеллектуальную систему (ИС) аналитической обработки цифрового сетевого контента (АОЦСК) в интересах эффективного обнаружения и противодействия нежелательной информации.

Элементы и уровни общей архитектуры интеллектуальной системы аналитической обработки цифрового сетевого контента

Понимая под *архитектурой* принципиальную организацию системы, воплощенную в её элементах, их взаимоотношениях друг с другом и со средой, а также принципы, направляющие её проектирование и эволюцию, остановимся на формулировке уровней (граней) общей архитектуры интеллектуальной системы аналитической обработки цифрового сетевого контента в интересах защиты от нежелательной информации.

Первый уровень (грань) общей архитектуры, отвечающий за *компетенции* интеллектуальной системы аналитической обработки цифрового сетевого контента, содержит:

(1) подуровень сбора и предварительной обработки данных о безопасности сетевых информационных объектов;

(2) подуровень оценивания смыслового содержания информации; подуровень выработки мер противодействия.

В рамках этих подуровней существуют элементы общей архитектуры, характеризующие *компоненты* интеллектуальной системы аналитической обработки цифрового сетевого контента. Этими компонентами являются:

(1) распределенные интеллектуальные сканеры сетевого цифрового контента, обеспечивающие оперативный, надежный и устойчивый к внешним воздействиям сбор и предварительную оценку информационных ресурсов сети Интернет и социальных сетей;

(2) компоненты многоаспектной оценки и категоризации смыслового наполнения информационных объектов (содержащий средства анализа текстовых, структурных, адресных, связностных, мультимедийных и др. признаков);

(3) компоненты обеспечения своевременности многоуровневого и многомодульного анализа информационных объектов на основе использования параллельных вычислений;

(4) компоненты устранения неполноты и противоречивости оценки и категоризации смыслового наполнения информационных объектов на основе использования методов обработки неполных, противоречивых и нечетких знаний;

(5) компоненты адаптации и переобучения системы анализа информационных объектов, в том числе в режиме эксплуатации;

(6) компоненты выработки и выбора мер противодействия нежелательной, сомнительной и вредоносной информации;

(7) компоненты реализации визуальных интерфейсов для выявления и противодействия нежелательной, сомнительной и вредоносной информации.

При этом в ИС АОЦСК объединены и совместно используются интеллектуальные модели и методы аналитической обработки сетевого цифрового контента в интересах защиты от нежелательной и вредоносной информации, распространяющейся в сети Интернет и социальных сетях.

Функциями и особенностями распределенных интеллектуальных сканеров является реализация децентрализованных режимов сбора и предобработки данных за счет наделения этих сканеров интеллектуальными свойствами.

Функциями и особенностями компонент многоаспектной оценки и категоризации смыслового наполнения информационных объектов является использование методов интеллектуального анализа данных и основанных на них классификаторов, которые учитывают достаточно большое количество признаков.

Компоненты обеспечения своевременности многоуровневого и многомодульного анализа информационных объектов обеспечивают реализацию комплекса взаимоувязанных классификаторов на программно-инструментальных платформах распределенной параллельной обработки данных.

Для устранения аспектов неполноты и противоречивости оценки и категоризации смыслового наполнения информационных объектов, которые свойственны этому процессу в силу самого

характера распространения вредоносной и нежелательной информации в сети Интернет и в социальных сетях, применяется усовершенствованные классификаторы информационных объектов.

Функциями и особенностями компонент адаптации и переобучения системы анализа информационных объектов, в том числе в режиме эксплуатации, является использование экспертных знаний и существующих списков классифицированных информационных объектов.

Компоненты выработки и выбора мер противодействия нежелательной, сомнительной и вредоносной информации отличаются от существующих аналогов использованием особенностей хранилища информации, построенного на основе файловой системы HDFS, а также применением механизма логического вывода, основанного на экспертных знаниях, использованием различных методик вычисления показателей (метрик) вредоносности и опасности цифрового контента и анализа рисков для последующего выбора соответствующих контрмер, а также сочетанием как методов автоматического вывода, так и визуального анализа результатов анализа для формирования мер противодействия.

Функциями и особенностями компонент реализации визуальных интерфейсов для выявления и противодействия нежелательной, сомнительной и вредоносной информации является то, что они ориентированы на использование нестандартных моделей визуализации, к числу которых следует отнести такие модели, как графы с глифами, карты деревьев, диаграммы Вороного и другие, а также модели виртуальной и дополненной реальности.

Второй уровень (грань) общей архитектуры, отвечающий за *интеграцию* в рамках интеллектуальной системы аналитической обработки цифрового сетевого контента, является *интеграционной платформой*, включающей четыре ключевых подуровня:

(1) подуровень аппаратно-программных средств интеллектуального анализа и принятия решений;

(2) подуровень хранилища информации о результатах интеллектуального анализа и принятия решений;

(3) подуровень аналитических инструментов и средств анализа, принятия решений и генерации отчетов;

(4) подуровень средств управления и настраиваемые интерфейсы с пользователями.

Интеграционная платформа является ядром системы, она реализует функции по интеграции и взаимодействию всех компонент, составляющих систему.

Информационная технология и система интеллектуальной аналитической обработки цифрового сетевого контента в интересах эффективного обнаружения, предупреждения и противодействия нежелательной информации отличаются тем, что в них интегрированы и объединены единой целевой установкой различные по своей направленности модели, методы, методики и программные приложения, основанные на применении интеллектуального анализа данных, обработки неполных, противоречивых и нечетких знаний, параллельных вычислений и визуализации данных.

Предлагаемые модели и методы учитывают специфику конкретных предметных областей, которые выбраны для оценки их эффективности. Это система родительского контроля цифрового контента в сети Интернет, система выявления и противодействия террористической и экстремистской активности в Интернете и социальных сетях, система выявления групповой вредоносной и нежелательной информационной активности (суицид, наркотики и др.) и выявления руководителей и членов соответствующих групп в социальных сетях.

Заключение

Таким образом, предложена общая архитектура интеллектуальной системы аналитической обработки цифрового сетевого контента в интересах защиты от нежелательной информации. Данная архитектура представляет собой совокупность элементов ИС АОЦСК, их взаимоотношения друг с другом и со средой. Это комплекс компонентов, средств, методов, методик, моделей, алгоритмов и программных прототипов интеллектуальной аналитической обработки цифрового сетевого контента в интересах эффективного обнаружения и противодействия нежелательной, сомнительной и вредоносной информации, распространяемой в сети Интернет и в социальных сетях.

Разработанные компоненты общей архитектуры, модели и методы найдут применение в предметных областях: системы родительского контроля в сети Интернет; системы выявления и противодействия террористической и экстремистской активности в Интернете и социальных сетях; системы анализа социальных сетей на выявление групповой вредоносной и нежелательной информационной активности.

Возможными потребителями научных и научно-технических результатов являются предприятия и организации телекоммуникационного сектора, силовые министерства и ведомства, административно-государственные структуры, в которых широко используются компьютерные системы и сети и предъявляются повышенные требования по защите от нежелательной и вредоносной информации. Кроме того, потребителями могут быть российские и иностранные коммерческие организации, занимающиеся разработкой и внедрением перспективных средств родительского контроля, фильтрации спама и аналитической обработки в социальных сетях, в частности, и в информационно-коммуникационных системах, в целом.

Возможным направлением предполагаемых дальнейших исследований может быть использование в качестве архитектурного решения концепции вычислений «в облаке» («computation in a cloud»), известную как «Безопасность как сервис» («Security as a service», SAAS). Тогда в качестве «облака» будет выступать back-end-сервер (система серверов) вместе с соответствующим программным и информационным обеспечением. Кроме того, в этом случае для одновременной обработки информационных объектов могут быть использованы элементы технологий, предназначенных для работы с Большими Данными.

Работа выполнена при финансовой поддержке гранта РНФ (проект № 18-11-00302) в СПИИРАН.

ЛИТЕРАТУРА

1. Baykan E., Henzinger M., Marian L., Beber I. Purely URL-based topic classification // *Proceedings of the 18th international conference on World wide web* (WWW '09). ACM, New York, NY, USA, 2009. P. 1109-1110.
2. Calado P., Cristo M., Moura E., Ziviani N., Ribeiro-Neto B., Goncalves M. A.. Combining link-based and content-based methods for web document classification // CIKM ’03: *Proceedings of the 12th International Conference on Information and Knowledge Management*, New York, NY, ACM Press. 2003. P. 394–401.
3. Dumais S., Chen H. Hierarchical classification of web content // *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval* (SIGIR '00). ACM, New York, NY, USA, 2000. P. 256-263.
4. Joachims T. Text categorization with support vector machines: learning with many relevant features // *Proceedings of ECML-98, 10th European Conference on Machine Learning*, Chemnitz, Germany, 1998. P. 137–142.
5. Kan M.-Y. Web page classification without the web page // WWW Alt.’04: Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, ACM, New York, NY, USA, 2004. P. 262—263.
6. Kan M.-Y., Thi H.O.N.. Fast web page classification using url features // CIKM ’05: *Proceedings of the 14th ACM international conference on Information and knowledge management,* ACM, New York, NY, USA, 2005. P. 325—326.
7. Базенков Н.И., Губанов Д.А. Обзор информационных систем анализа социальных сетей // Управление большими системами. Выпуск 41. М.: ИПУ РАН, 2013. С. 357-394. cyberleninka.ru/article/n/obzor-informatsionnyh-sistem-analiza-sotsialnyh-setey.pdf
8. Котенко И.В., Чечулин А.А., Комашинский Д.В. Автоматизированное категорирование веб-сайтов для блокировки веб-страниц с неприемлемым содержимым // Проблемы информационной безопасности. Компьютерные системы, № 2, 2015. С. 69-79.
9. Котенко И.В., Юсупов Р.М. Информационные технологии для борьбы с терроризмом // Защита информации. INCIDE № 2, 2009. С. 74-79.
10. Новожилов Д.А., Чечулин А.А., Котенко И.В. Улучшение категорирования веб-сайтов для блокировки неприемлемого содержимого на основе анализа статистики html-тэгов // Информационно-управляющие системы, № 6, 2016. С. 65-73.
11. Рукавицын А.Н. Разработка модели классификации веб-страниц с использованием методов интеллектуального анализа данных // Известия СПбГЭТУ «ЛЭТИ» No 4/2016. С. 12-20.
12. Kotenko I., Chechulin A., Komashinsky D. Evaluation of Text Classification Techniques for Inappropriate Web Content Blocking // *The IEEE 8th International Conference on «Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*» (IDAACS'2015). Proceedings. Warsaw, Poland, September 24-26, 2015. P. 412-417.
13. Kotenko I., Chechulin A., Komashinsky D. Categorisation of web pages for protection against inappropriate content in the internet // *International Journal of Internet Protocol Technology* (IJIPT), Vol. 10, No. 1, 2017. P. 61-71.

14. Котенко И.В., Саенко И.Б. SIEM-системы для управления информацией и событиями безопасности // Защита информации. Инсайд. 2012. № 5 (47). С.54-65.
15. Котенко И.В., Саенко И.Б. Архитектура системы интеллектуальных сервисов защиты информации в критически важных инфраструктурах // Труды СПИИРАН. № 1 (24). 2013. С.21-40.

I.V. Kotenko, I.B. Parashchuk (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), St. Petersburg)

General architecture of intelligent system for analytical processing of digital network content in order to protect against unwanted information

Abstract. The report examines the elements and levels of the general architecture of an intelligent system for analyzing digital network content, their functions and features. This architecture is a set of tools, methods, techniques, models, algorithms and software prototypes of intelligent analytical processing of digital network content in the interests of effective detection and countering of unwanted, questionable and harmful information distributed on the Internet and social networks. The components of the general architecture will find application in the systems of parental control, systems for identifying and countering terrorist and extremist activity on the Internet and social networks, as well as in social networks analysis systems, to identify group harmful and undesirable information activity.

ПОКАЗАТЕЛИ КАЧЕСТВА СИСТЕМЫ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В УПРАВЛЕНИИ И ОЦЕНИВАНИЕ ЭФФЕКТИВНОСТИ ЕЕ ФУНКЦИОНИРОВАНИЯ

И. Б. Паращук[1]
(*Санкт-Петербургский институт информатики и автоматизации
Российской академии наук (СПИИРАН),
199178, Россия, Санкт-Петербург, 14-я линия В.О., д. 39,
(812) 328-71-81, parashchuk@comsec.spb.ru*)

Аннотация

Содержание работы составляют вопросы, посвященные обоснованию актуальности, а также анализу условий и методов оценивания эффективности функционирования систем обеспечения безопасности информационных технологий, выбору и обоснованию показателей качества и метода формирования обобщенного показателя эффективности функционирования таких систем, разработке методики и алгоритмов векторного динамического оценивания эффективности функционирования систем обеспечения безопасности информационных технологий в управлении.

Выработаны организационно-технические предложения по совершенствованию структуры и алгоритмов работы подсистемы управления в рамках систем обеспечения безопасности информационных технологий.

Ключевые слова: эффективность, показатель, безопасность, технология, качество.

Введение

Анализ современных условий, в которых осуществляется функционирование информационных технологий в управлении, позволяет выделить целый ряд противоречий, обусловленных ужесточением требований к качеству процессов обеспечения безопасности информации, обеспечения безопасности информационных технологий в управлении в рамках обмена всеми видами информации и постоянно растущими возможностями нарушителей по реализации угроз безопасности информации, включая компьютерные атаки [1-5].

Основой методологии построения систем обеспечения безопасности информационных технологий (СОБИТ) являются методы оценивания эффективности их функционирования [6, 7]. Но необходимый в существующих методиках сбор статистики о состоянии элементов СОБИТ, требует значительного времени и снижает точность и оперативность анализа, следовательно – повышает время реакции подсистемы автоматизированного управления СОБИТ на реализуемые нарушителем угрозы и вводимые управляющие воздействия.

Цель работы состоит в повышении точности и оперативности анализа эффективности функционирования СОБИТ в условиях различного вида воздействий на основе стохастического оценивания показателей качества систем такого класса. Предложены методика и алгоритмы векторного динамического оценивания эффективности функционирования СОБИТ, использующие вероятностное описание модели СОБИТ как управляемой системы, состоятельные системы показателей качества (СПК) в виде отклонений показателей качества системы обеспечения безопасности информации от предъявляемых к ним требований, марковские модели этих показателей качества, учитывающие динамику процесса функционирования системы обеспечения безопасности информации, методы стохастического оценивания значений отклонений показателей качества, а также учитывающие вклад переходных процессов, возникающих в результате реализации различного вида угроз и управляющих воздействий на СОБИТ, в общую эффективность их функционирования.

Система показателей качества и методика оценивания эффективности СОБИТ

Одним из основных этапов разработки методики ВДОЭФ СОБИТ является этап обоснования иерархически взаимоувязанной СПК функционирования управляемых систем такого класса. Проведенный анализ различных методов формирования СПК сложных систем показал целе-

[1] Доктор технических наук, профессор, ведущий научный сотрудник.

сообразность использования, в качестве прототипа, подхода к формированию СПК, предложенного в работах [8, 9], основанного на математических методах теории функциональной и параметрической декомпозиции. Однако, предложенная в данных работах СПК не применима для СОБИТ и носит неполный характер, поскольку в ней невозможно учесть числовые характеристики отклонений оптимизируемых параметров СОБИТ от предъявляемых к ним требований, что делает затруднительным использование такой СПК при анализе качества и эффективности управляемых СОБИТ с ограниченным ресурсом, где отклонения параметров СОБИТ от требований к ним являются дополнительными характеристиками качества управления обеспечением безопасности информации. Предложена СПК функционирования управляемой СОБИТ, которая, в отличие от существующих, содержит показатели качества в виде текущих отклонений параметров СОБИТ от требований к ним. Так, текущий векторный показатель качества функционирования СОБИТ, характеризует такие наиболее важные свойства системы как доступность, целостность, конфиденциальность информации и ресурсопотребление на реализацию процесса обеспечения безопасности информации и имеет вид:

$$\Delta\vec{Y}_{\text{ф}}(k)=[\Delta\bar{t}_{\text{дост}}^{\,усл/\,угр}(k);\ \Delta K_{\text{п дост}}(k);\ \Delta K_{\text{иск}}(k);\ \Delta K_{\text{конф инф}}(k);\ \Delta\vec{З}_{\text{ф}}(k)]^{\text{т}}, \quad (1)$$

где доступность характеризует $\Delta\bar{t}_{\text{дост}}^{\,усл/\,угр}(k)$ – отклонения времени доступа легальных (авторизированных) пользователей к защищаемому информационному ресурсу на *k*-ом шаге функционирования СОБИТ; целостность (достоверность и неискаженность информации, несмотря на наличие угроз и уязвимостей) характеризуют $\Delta K_{\text{п дост}}(k)$ – отклонения коэффициента потери достоверности информации и $\Delta K_{\text{иск}}(k)$ – отклонения коэффициента искажений на *k*-ом шаге функционирования СОБИТ; конфиденциальность (способность СОБИТ сохранять информацию в тайне от субъектов, не имеющих полномочий на доступ к ней) характеризует $\Delta K_{\text{конф инф}}(\text{k})$ – отклонения коэффициента конфиденциальности информации на *k*-ом шаге функционирования СОБИТ. Помимо этого, выражение (1) содержит отклонения вектора затрат ресурсов $\Delta\vec{З}_{\text{ф}}(k)$ на построение СОБИТ и реализацию процесса ее функционирования.

Выражение (1) представляет собой формализованную запись глобальной СПК, характеризующей общую задачу, стоящую перед СОБИТ. Система показателей качества СОБИТ включает глобальную СПК и локальные СПК процесса обеспечения безопасности информации, процесса управления обеспечением безопасности информации, комплекса средств обеспечения безопасности информации и подсистемы управления СОБИТ. Ведущей среди локальных СПК выступает локальная СПК процесса обеспечения безопасности информации, компоненты которой аналогичны, во многом, компонентам глобальной СПК СОБИТ, имеющая вид

$$\Delta\vec{Y}_{\text{оби}}(k)=[\Delta\vec{Y}_{\text{дост}}(k);\Delta\vec{Y}_{\text{непр}}(k);\Delta\vec{Y}_{\text{цел}}(k);\Delta\vec{Y}_{\text{конф}}(k);\Delta\vec{З}_{\text{зи}}(k)]^{\text{т}}, \quad (2)$$

и включающая в себя векторы показателей доступности легальных (авторизированных) пользователей информационных технологий к защищаемому информационному ресурсу, непрерывности процесса обеспечения безопасности информации, целостности информации при реализации этого процесса, конфиденциальности информации и вектор показателе затрат ресурсов на реализацию процесса обеспечения безопасности информации.

Функции управления процессом обеспечения безопасности информации обеспечивают собственно процесс защиты информации, поэтому локальная СПК управления процессом обеспечения безопасности информации имеет более низкий уровень иерархии по сравнению с локальной СПК данного процесса и включает: отклонения длительности цикла управления параметрами процесса обеспечения безопасности информации при нарушении нормальных условий функционирования на *k*-ом шаге; отклонения вектора приращений значений показателей качества процесса обеспечения безопасности информации (определяющего точность управления этим процессом) на *k*-ом шаге, обусловленных ошибками в контуре управления; отклонения времени доступности потенциального нарушителя к сигналам и командам управления процессом обеспечения безопасности информации на *k*-ом шаге и отклонения вектора затрат ресурсов управления на *k*-ом шаге реализации процесса обеспечения безопасности информации

$$\Delta\vec{Y}_{\text{у}}(k)=[\Delta T_{\text{цу}}(k);\Delta\vec{Y}_{\text{оби}}^{\text{у}}(k);\Delta\bar{t}_{\text{ду}}(k);\Delta\vec{З}_{\text{у}}(k)]^{\text{т}}\ . \quad (3)$$

Материальной основой процесса обеспечения безопасности информации и управления этом процессом являются два ключевых элемента СОБИТ: комплекс средств обеспечения безопасности информации и подсистема управления, иерархия которых определяется иерархией осуществляемых ими процессов и иерархией управления. Компонентами локальной СПК комплекса средств обеспечения безопасности информации являются векторы отклонений от требуемых значений показателей производительности, устойчивости, разведзащищенности, достоверности функционирования, масштабируемости, совместимости, мультисервисности, безопасности и ресурсопотребления данного комплекса на *k*-ом шаге функционирования СОБИТ:

$$\Delta\vec{Y}_{\text{ксоби}}(k)=[\Delta\vec{Y}_{\text{пр ксоби}}(k);\Delta\vec{Y}_{\text{уст ксоби}}(k);\Delta\vec{Y}_{\text{рз ксоби}}(k);\Delta\vec{Y}_{\text{достов ксоби}}(k);$$
$$\Delta\vec{Y}_{\text{масшт ксоби}}(k);\Delta\vec{Y}_{\text{совмест ксоби}}(k);\Delta\vec{Y}_{\text{м/серв ксоби}}(k);\Delta\vec{Y}_{\text{безоп ксоби}}(k);\Delta\vec{З}_{\text{ксоби}}(k)]^{\text{т}}. \quad (4)$$

Последней в иерархии систем показателей качества СОБИТ является СПК подсистемы управления СОБИТ, имеющая вид

$$\Delta\vec{Y}_{\text{пу}}(k)=[\Delta V^{угр}_{\text{тех пу}}(k);\Delta T_{\text{цу пу}}(k);\Delta\vec{Y}^{\text{у}}_{\text{пу}}(k);\Delta\bar{t}_{\text{бсф пу}}(k);\Delta\bar{t}_{\text{ду пу}}(k);\Delta\vec{З}_{\text{пу}}(k)]^{\text{т}} \quad (5)$$

и включающая отклонения показателей качества производительности подсистемы управления СОБИТ, оперативности (своевременности) управления, достоверности, устойчивости, непрерывности и скрытности управления комплексом средств обеспечения безопасности информации, а также затраты ресурсов на реализацию подсистемы управления на *k*-ом шаге функционирования СОБИТ.

Таким образом, предложенная СПК СОБИТ, развитая на случай динамического анализа качества и эффективности функционирования систем обеспечения безопасности информации такого класса, обладая полнотой и безизбыточностью, вместе с тем, позволяет расширить диапазон исследуемых характеристик управляемой СОБИТ. Такой подход справедлив для сложных систем с жестким замкнутым контуром управления, у которых имеется достаточный ресурс для поддержания показателей качества в области требуемых значений. Кроме того, в работе И.Е.Казакова и В.М.Артемьева [10] доказано положение о нормализации плотностей распределения вероятностей ошибок управления (значений отклонений показателей качества от требуемых) в системах управления с замкнутой обратной связью, что позволяет унифицировать алгоритмы оценивания различных по классу частных показателей эффективности функционирования управляемых СОБИТ. Понимая под эффективностью функционирования СОБИТ степень соответствия показателей качества функционирования системы требованиям, предъявляемым со стороны пользователей, с учетом особенностей разработанной СПК СОБИТ, случайного характера изменения большинства показателей качества СОБИТ и возможности решения в рамках выбранного вероятностного подхода основных проблем многокритериального динамического оценивания эффективности, текущие частные показатели эффективности функционирования СОБИТ и обобщенный показатель эффективности функционирования СОБИТ сформулированы с применением аппарата условных вероятностей и методов вероятностной скаляризации в виде совместной вероятности выполнения требований к качеству процесса обеспечения безопасности информации, процесса управления обеспечением безопасности информации, комплекса средств обеспечения безопасности информации и подсистемы управления на *k*-ом шаге функционирования СОБИТ с учетом выражений (1) – (5).

Для методики, учитывающей вероятностно-временной механизм изменения показателей качества СОБИТ на *k*-ом шаге ее функционирования, обобщенный показатель эффективности функционирования СОБИТ записывается в виде:

$$\begin{aligned}P_{\text{вып ф}}(k)&=P(\Delta\vec{Y}_{\text{ф СОБИТ}}(k)\le\Delta\vec{Y}^{\text{тр}}_{\text{ф СОБИТ}})=P[(\Delta\vec{Y}_{\text{ф}}(k)\le\Delta\vec{Y}^{\text{тр}}_{\text{ф}})/(\Delta\vec{Y}_{\text{СОБИТ}}(k)\le\Delta\vec{Y}^{\text{тр}}_{\text{СОБИТ}})=\\&=P_{\text{вып оби}}(k)[(\Delta\vec{Y}_{\text{оби}}(k)\le\Delta\vec{Y}^{\text{тр}}_{\text{оби}})/(\Delta\vec{Y}_{\text{у}}(k)\le\Delta\vec{Y}^{\text{тр}}_{\text{у}})\bigcap(\Delta\vec{Y}_{\text{ксоби}}(k)\le\Delta\vec{Y}^{\text{тр}}_{\text{ксоби}})\bigcap\\&\bigcap(\Delta\vec{Y}_{\text{пу}}(k)\le\Delta\vec{Y}^{\text{тр}}_{\text{пу}})]\times P_{\text{вып у}}(k)[(\Delta\vec{Y}_{\text{у}}(k)\le\Delta\vec{Y}^{\text{тр}}_{\text{у}})/(\Delta\vec{Y}_{\text{ксоби}}(k)\le\Delta\vec{Y}^{\text{тр}}_{\text{ксоби}})\bigcap\\&\bigcap(\Delta\vec{Y}_{\text{пу}}(k)\le\Delta\vec{Y}^{\text{тр}}_{\text{пу}})]\times P_{\text{вып ксоби}}(k)[(\Delta\vec{Y}_{\text{ксоби}}(k)\le\Delta\vec{Y}^{\text{тр}}_{\text{ксоби}})/\\&/(\Delta\vec{Y}_{\text{пу}}(k)\le\Delta\vec{Y}^{\text{тр}}_{\text{пу}})]\times P_{\text{вып пу}}(k)[(\Delta\vec{Y}_{\text{пу}}(k)\le\Delta\vec{Y}^{\text{тр}}_{\text{пу}})],\end{aligned} \quad (6)$$

где $\Delta\vec{Y}_{\text{ф}}(k),\Delta\vec{Y}_{\text{СОБИТ}}(k),\Delta\vec{Y}^{\text{тр}}_{\text{ф}},\Delta\vec{Y}^{\text{тр}}_{\text{СОБИТ}}$ – вектора показателей качества (в виде отклонений параметров от требований) функционирования СОБИТ на *k*-ом шаге и требования к ним; $P_{\text{вып оби}}(k),P_{\text{вып у}}(k),P_{\text{вып ксоби}}(k)$ – условные вероятности выполнения требований к качеству

процесса обеспечения безопасности информации, управления данным процессом и комплекса средств обеспечения безопасности информации на *k*-ом шаге функционирования СОБИТ, определяемые при условии выполнения требований к показателям качества подсистемы управления СОБИТ; $P_{\text{вып пу}}(k)$ – безусловная вероятность выполнения требований к показателям качества подсистемы управления на *k*-ом шаге функционирования СОБИТ.

С учетом способа формирования частных и обобщенного (6) показателей эффективности функционирования СОБИТ, методика векторного динамического оценивания эффективности на каждом *k*-ом шаге функционирования включает ряд этапов:

Этап 1 – сбор статистических данных о значениях отклонений показателей качества СОБИТ $\Delta\vec{Y}_j(k) = f(\Delta y_1(k); \Delta y_2(k); \dots \Delta y_i(k); \dots \Delta y_N(k))$ за период наблюдения (на *k*-ом шаге функционирования СОБИТ):

$$\vec{z}_{\Delta\vec{Y}_j}(k) = \{ z_{\Delta y_1}(k); z_{\Delta y_2}(k); \dots z_{\Delta y_i}(k); \dots z_{\Delta y_N}(k)\}^{\text{т}};$$

Этап 2 – оптимальная (по критерию минимального среднего квадратичного отклонения) фильтрация значений отклонений показателей качества СОБИТ (условных средних) наблюдаемого процесса:

$$\Delta\hat{y}_i(k) = \text{M}\{\Delta y_i(k) / z_{\Delta y_i}(k)\}, \quad i = \overline{1, N};$$

Этап 3 – формирование оценочного значения частного (векторного) показателей качества СОБИТ ($\Delta\hat{\vec{Y}}_j(k)$) на основе оценочных значений его компонент:

$$\Delta\hat{\vec{Y}}_j(k) = f(\Delta\hat{y}_1(k); \Delta\hat{y}_2(k); \dots \Delta\hat{y}_i(k); \dots \Delta\hat{y}_N(k));$$

Этап 4 – идентификация параметров условной по наблюдениям плотности распределения вероятностей значений отклонений частного показателя качества СОБИТ $\hat{\varpi}(\Delta\vec{Y}_j(k))$ с использованием оценок его моментов $\text{M}\{\Delta\vec{Y}_j(k) / z_{\Delta\vec{Y}_j}(k)\}$, $\text{D}_{\Delta\vec{Y}_j}(k)$;

Этап 5 – Определение каждого частного показателя эффективности функционирования СОБИТ на *k*-ом шаге

$$\hat{P}_{\text{вып } j}(\Delta\hat{\vec{Y}}_j(k) \le \Delta\vec{Y}_j^{\text{тр}}) = \int_0^{\Delta\vec{Y}_j^{\text{тр}}} \hat{\varpi}(\Delta\vec{Y}_j(k)/z_{\Delta\vec{Y}_j}(k))\, d\Delta\vec{Y}_j;$$

Этап 6 – Определение обобщенного показателя эффективности функционирования СОБИТ для *k*-ого шага в соответствии с выражением (6)

$$\hat{P}_{\text{вып ф}}(k) = P(\Delta\hat{\vec{Y}}_{\text{ф СОБИТ}}(k) \le \Delta\vec{Y}^{\text{тр}}_{\text{ф СОБИТ}}).$$

Проверена конструктивность разработанных методики и алгоритмов ВДОЭФ СОБИТ, проверена применимость полученного инструмента анализа эффективности для оценивания СОБИТ различной структуры в целом в динамике их функционирования, а также разработаны предложения по совершенствованию подсистемы управления СОБИТ.

С целью повышения эффективности функционирования СОБИТ предложено создание единого контура управления перераспределением входящего информационного потока между иерархическими ступенями общей системы обеспечения безопасности информации с учетом поступающей информации о текущем качестве и эффективности элементов и подсистем общей системы защиты.

Заключение

Таким образом, на основе анализа возможностей новых технологий защиты в современных и перспективных информационных технологиях для управления, сделано теоретическое обобщение существующих методов анализа качества систем обеспечения безопасности информационных технологий, разработана методика векторного динамического оценивания эффективности функционирования таких систем, учитывающая вероятностный и динамический характер процесса функционирования системы в условиях различного вида воздействий.

Предложена методика векторного динамического оценивания эффективности функционирования систем обеспечения безопасности информационных технологий, использующая вероятностное описание модели управляемой системы; состоятельные показатели качества управляемых систем обеспечения безопасности информационных технологий в виде отклонений от предъявляемых к ним требований; марковские модели этих показателей качества; методы стохастического оценивания значений отклонений показателей качества, учитывающие вклад переходных процессов в общую эффективность функционирования систем обеспечения безопасности информационных технологий.

Выработаны организационно-технические предложения по совершенствованию структуры и алгоритмов работы подсистемы управления систем обеспечения безопасности информационных технологий.

Возможными направлениями предполагаемых дальнейших исследований могут быть синтез состоятельных систем показателей качества и формулировка этапов анализа эффективности систем такого класса в условиях недостоверности, неполноты, нечеткости и противоречивости исходных данных.

Работа выполнена при финансовой поддержке гранта РНФ (проект № 18-11-00302) в СПИИРАН.

ЛИТЕРАТУРА

1. Kotenko I.V. Active Vulnerability Assessment of Computer Networks by Simulation of Complex Remote Attacks. *Proceedings of the International Conference on Computer Networks and Mobile Computing*. ICCNMC-03 (IEEE Computer Society). 2013. P. 40-47.
2. Kotenko I.V., Saenko I.B., Ageev S.A., Kopchak Y.M. Abnormal Traffic Detection in networks of the Internet of things based on fuzzy logical inference *XVIII International Conference on Soft Computing and Measurements* (*SCM'2015*) *IEEE Xplore.* 2015. P. 5–8.
3. Kotenko I.V., Chechulin A.A., Novikova E.S. Attack Modelling and Security Evaluation for Security Information and Event Management. *Proceedings of the International Conference on Security and Cryptography* (SECRYPT 2012). 2012. P. 391-394.
4. Kotenko I.V. Agent-Based Modeling and Simulation of Cyber-Warfare between Malefactors and Security Agents in Internet. *Proceedings of the 19th European Simulation Multiconference «Simulation in wider Europe»* (ESM'05). 2005. P. 533-543.
5. Kotenko I., Konovalov A., Shorov A. Agent-based simulation of cooperative defence against botnets // Concurrency Computation Practice and Experience. 2012. Vol.24. № 6. P.573-588.
6. Бусленко Н.П. Моделирование сложных систем. М.: Наука, 1978. 400 с.
7. Петухов Г.Б. Основы теории эффективности целенаправленных процессов. Часть I. Методология, методы, модели. М.: МО СССР, 1989. 660 с.
8. Терентьев В.М., Санин Ю.В. Анализ эффективности функционирования автоматизированных сетей многоканальной радиосвязи. СПб.: ВАС, 1992. 80 с.
9. Паращук И.Б., Башкирцев А.С. К вопросу обоснования систем показателей качества процессов принятия решения и поддержки принятия решения в интересах управления информационными сетями. // Информация и космос. № 2, 2016. С. 65-71.
10. Казаков И.Е., Артемьев В.М. Оптимизация динамических систем случайной структуры. М.: Наука, 1980. 384с.

I.B. Parashchuk (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), St. Petersburg)

Indicators of the quality of the information technology security systems and its effectiveness evaluation

Abstract. The content of the work is the issues devoted to the substantiation of relevance, as well as the analysis of the conditions and methods for assessing the effectiveness of the functioning of security systems for information technology, the selection and justification of quality indicators and the method of forming a generalized indicator of the effectiveness of such systems, the development of methodologies and algorithms for vector dynamic estimation of the effectiveness of the functioning of information security management systems. Organizational and technical proposals for improving the structure and algorithms of the control subsystem within the information technology security systems are developed.

ПРОГРАММНОЕ СРЕДСТВО РАСПРЕДЕЛЕННОГО ВЫПОЛНЕНИЯ ПРОЦЕДУР МАШИННОГО ОБУЧЕНИЯ ДЛЯ ЦЕЛЕЙ СЕТЕВОЙ БЕЗОПАСНОСТИ

А. Г. Кушнеревич[1], **А. А. Браницкий**[2]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), С.-Петербург, 14-я линия В.О., д. 39, kushnerevich@comsec.spb.ru, branitskiy@comsec.spb.ru*)

Аннотация

Одной из наиболее популярных сегодня систем распределенной обработки больших данных является Spark и ее библиотека для процедур машинного обучения MLlib. В работе проводится оценка возможности и эффективности использования этой платформы для взаимодействия с алгоритмами машинного обучения (деревом решений, наивным байесовским классификатором, методом опорных векторов), которые могут применяться в том числе для задач защиты от информации.

Ключевые слова: Spark, MLlib, метод опорных векторов, деревья решений, наивный байесовский классификатор.

Введение

В современном информационном мире, ежедневно генерирующем колоссальные объемы данных, остро стоит проблема своевременной обработки информации. Эта необходимость обусловлена перспективами повышения качества жизни людей. Сегодня основными сферами применения технологий обработки и хранения больших данных являются бизнес и медицина. В сфере розничной интернет-торговли такие подходы применяются повсеместно: зная информацию о вас, моделях вашего поведения в сети, можно выявлять ваши ближайшие потребности и предлагать соответствующие продукты в рекламных блоках, которые есть практически на любом сайте. В медицинской сфере на основе анализов человека и накопленных знаний технологии больших данных позволяют ставить диагнозы на ранних стадиях и составлять прогнозы течения болезней человека, что иногда не под силу даже самым квалифицированным специалистам, тем самым спасая несчетное количество жизней.

Одной из тем технологий больших данных является машинное обучение. С развитием алгоритмов машинного обучения и ростом вычислительных мощностей оборудования стало возможным улучшение процессов выявления закономерностей в данных. Неся в себе существенную полезность для бизнеса, алгоритмы машинного обучения также обладают высокими запросами к оборудованию. Наличие больших данных для обучения алгоритмов требует ускоренной их обработки, чтобы соответствовать приемлемым в условиях производства срокам обучения.

Распространенными способами горизонтального увеличения вычислительных мощностей оборудования является применение подходов *MapReduce (Hadoop), RDD (Resilient Distribited Dataset, Spark)* [1].

Подход *MapReduce* позволяет распределить вычисления между узлами кластера для их параллельного выполнения за счет того, что все операции над данными в подходе делятся на два глобальных типа: *Map* и *Reduce*. Различие типов состоит в том, что операции *Map* выполняются над независимыми блоками данных и не требуют знания результатов обработки других блоков, в то время как операции *Reduce* требуют результаты обработки всех либо многих блоков, чтобы получить результат либо независимую часть результата [2].

В подходе RDD распределение вычислений происходит схожим образом (Spark может использовать YARN – менеджер ресурсов Hadoop). Все данные делятся на блоки и составляют специальные наборы данных RDD. Кроме того, RDD можно организовывать в потоки DStream с помощью библиотеки Streaming. Это позволяет говорить о возможности ведения вычислений над данными, поступающими в режиме реального времени. Как и в подходе MapReduce, в подходе RDD существуют операции типа *Map* и *Reduce*. Однако они представлены более высокоуровневыми операторами: так, к *Map* можно отнести операторы *map, flatMap, filter, foreach* и

[1] Младший научный сотрудник.
[2] Младший научный сотрудник.

др.; к операторам *Reduce* – *reduce*, *reduceByKey*, *countByKey*, *countByValue* и др. Кроме того, Spark реализует подход ленивых вычислений, в результате чего все операторы делятся на трансформации и действия. Операторы трансформации переводят один RDD в другой RDD, при этом не запуская вычислений, – обычно это какие-либо операции промежуточного вычисления, результат выполнения которых не представляет интереса. Поэтому с целью повышения эффективности распараллеливания вычисления могут запускать только операторы-действия. Операторы-действия преобразуют распределенные наборы данных RDD к локальным наборам данных (например, list или dict в языке Python). Как правило, оператору действия предшествует несколько операторов трансформации. В конечном счете формируется цепочка операторов, зная которую менеджер ресурсов может повысить эффективность работы задачи. Примером такой цепочки может являться представленная на рисунке 1 последовательность операторов [3].

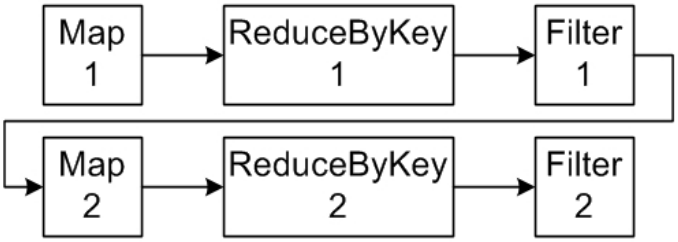


Рис. 1. RDD-цепочка операторов

Поскольку очевидным является тот факт, что подход RDD является более гибким и перспективным с точки зрения вычислений в реальном времени, исследование будет направлено на оценку возможностей и эффективности выполнения процедур машинного обучения именно на этой платформе.

Математические основы

Постановка задачи

В статье решается задача классификации объектов. Она формулируется следующим образом. Пусть дано множество пар, состоящее из признаковых описаний классифицируемых объектов в виде вектора и приписанной ему метки класса:

$$P = \{(x_i, c_i)\},\ i = 1,\ldots, M,$$

где x_i – вектор признакового описания i-го классифицируемого объекта; c_i – приписанная i-му объекту метка класса.

Требуется разработать алгоритм R, который позволит аппроксимировать множество P на основе имеющейся информации о векторах $\{x_i\}$: $\#\{x_i \mid R(x_i) \neq c_i\} \rightarrow \min$. Для решения этой задачи в статье используются методы машинного обучения, которые известны своей способностью выявлять скрытые закономерности в анализируемых данных.

Модели машинного обучения

Метод (машина) опорных векторов (Support Vector Machine, SVM) заключается в построении разделяющей гиперплоскости, обладающей свойством равноудаленности от ближайших к ней объектов из разных классов [4]. Модель представляется следующим образом:

$$F(z) = sign\left(-b + \sum_{i=1}^{M_s} w_i x_i^T z\right),$$

где w_i – весовые коэффициенты, представляющие собой произведение ненулевых множителей Лагранжа и желаемых выходных значений; x_i – опорные векторы ($i = 1, .. , M_s$); b – параметр смещения. В данной формуле предполагается, что обучающая выборка может быть линейно разделена. В противном случае необходимо применение специальных преобразований – ядер.

Дерево решений (Decision Tree, DT) – иерархическая структура, содержащая в качестве нетерминальных узлов числовые признаки и предикаты, вычисляемые на этих признаках, а в качестве терминальных узлов – метки классов. По мере спуска вниз по дереву и в зависимости от

истинности предиката для компонента наблюдаемого вектора выбирается один из двух путей в дереве решений:

$$F(z) = R(T, z)$$

$$R(T, z) = \begin{cases} c, \text{ если } T \text{ содержит только терминальный узел} \\ R(T^{(L)}, z), \text{ если } P^{(T)}(z) \\ R(T^{(R)}, z), \text{ если не } P^{(T)}(z) \end{cases},$$

где *T(L)* и *T(R)* – левое и правое поддеревья *T*; *c* – метка класса, *P(T)* – предикат, расположенный в корне дерева *T*. Данное определение подразумевает рекурсивное вычисление метки класса посредством обрезания дерева на одну из двух частей.

Наивный байесовский классификатор (Naïve Bayes, NB) представляет собой семейство вероятностных классификаторов, основанных на применении теоремы Байеса [5]. Так называемая «наивность» заключается в предположении о строгой независимости каждой пары признаков. Может применяться как для бинарной классификации, так и для многоклассовой.

$$F(z) = \arg\max_{c \in C}[P(z \mid c)P(c)],$$

где *P(z|c)* – вероятность встретить запись *z* среди всех объектов, принадлежащих классу *c*; *P(c)* – безусловная вероятность появления записи класса *c* в наборе данных.

Экспериментальные исследования

В качестве тестовой машины был использован сервер на базе гипервизора ESXi 6.0. Виртуальной машине было выделено 8 потоков ЦПУ по 2.1 ГГц.

Тестовый набор данных [6] состоял из 3082456 объектов общим объемом 1115,1 МБ. Набор представлял собой файл формата CSV и содержал 106 признаков, не включая метки класса и идентификатора объекта. Требовалось определить, к какому из двух классов относится объект: к нормальному трафику или к сканированию (syn или udp). С целью выявления преимуществ быстродействия Spark в случае параллелелизации алгоритмов машинного обучения, были проведены экспериментальные исследования в локальном режиме Python и с использованием библиотеки MLlib (Spark). Тестовый набор данных был разделен в пропорции 9:1, 9 частей приходилось на обучение и 1 часть – на тестирование модели. Результаты экспериментов приведены в таблице 1.

Таблица 1

Результаты тестирования алгоритмов машинного обучения

Режим выполнения	Decision Tree		SVM		Naïve Bayes	
	Время	Точность	Время	Точность	Время	Точность
параллельный	299,2 с	100%	1623,4	100%	275,2	69,6%
локальный	1974,12	100%	53928,8	100%	1765,2	69,5%

Как видно из таблицы 1, при равной точности скорость параллельного выполнения обучения и тестирования в среднем превосходит в 6,5 раз этот же показатель, вычисленный для локального режима. Для машины опорных векторов она и вовсе оказалась выше в 33 раза. Это обусловлено использованием восьми параллельных потоков вместо одного в локальном режиме, который не позволяет использовать большее число потоков.

Заключение

Таким образом, по результатам вычислений в параллельном режиме Spark и локальном режиме языка Python можно сделать вывод о том, что библиотека MLlib из пакета Spark не подвержена ограничениям GIL, в отличие от интерпретатора Python. Также это свидетельствует о том, что такие ресурсоемкие задачи машинного обучения, как задачи классификации, кластеризации и регрессии, могут быть эффективно выполнены средствами Spark и MLlib.

Работа выполнена при финансовой поддержке гранта РНФ (проект № 18-11-00302) в СПИИРАН.

ЛИТЕРАТУРА

2. Kotenko I, Saenko I., and Kushnerevich A. Parallel big data processing system for security monitoring in Internet of Things networks // Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA). 2017. Vol. 8. № 4. P. 60-74.
3. Kotenko I, Fedorchenko A., Saenko I., Kushnerevich A.. Parallelization of Security Event Correlation Based on Accounting of Event Type Links // International Conference on Parallel, Distributed and Network-based Processing (PDP). IEEE, 2018. P. 462-469.
4. Котенко И.В., Федорченко А.В., Саенко И.Б., Кушнеревич А.Г. Технологии больших данных для корреляции событий безопасности на основе учета типов связей // Вопросы кибербезопасности, 2017. № 5(23). С. 2-16
5. Cortes C., Vapnik V. Support-vector networks // Machine learning. 1995. Т. 20. №. 3. С. 273-297.
6. Zhang H. The optimality of naive bayes // Seventeenth International Florida Artificial Intelligence Research Society Conference. 2004. P. 562–567.
7. Branitskiy A., Kotenko I. Network Anomaly Detection Based on an Ensemble of Adaptive Binary Classifiers // International Conference on Mathematical Methods, Models, and Architectures for Computer Network Security. 2017 P. 143-157.

A.G. Kushnerevich, A.A. Branitskiy (SPIIRAS, Saint-Petersburg)

Software for Distributed Machine Learning procedures for Network Security Purposes

Today Spark and MLlib (its library for machine learning procedures) are among the most popular distributed data processing systems for Big Data. In this paper, we will evaluate the possibility and efficiency of using this platform for interaction with machine learning algorithms such as the decision tree, the naive Bayesian classifier, the support vector machine.

ПОДДЕРЖКА ПРИНЯТИЯ РЕШЕНИЙ ПО ПРОТИВОДЕЙСТВИЮ НЕЖЕЛАТЕЛЬНОЙ ИНФОРМАЦИИ

Л. А. Виткова[1], **Е. В. Дойникова**[2]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН),*
199178, Россия, С.-Петербург, 14-я линия В.О., д. 39,
(812) 328-71-81, iskinlidia@gmail.com, doynikova@comsec.spb.ru)

Аннотация

В настоящее время все очевиднее становится проблема информационно-психологического ущерба от нежелательной информации, что подтверждает значимость проблемы противодействия нежелательной информации. Современные мероприятия в области защиты от нежелательной информации в Российской Федерации в основном направлены на выработку нормативно-правовых мер и носят предупредительный характер. При этом, существующие нормативные акты не имеют публичных пояснений и комментариев по отнесению информации к нежелательной. На сегодняшний день не существует полноценного инструмента автоматизированного поиска и защиты от нежелательной информации в сети Интернет. Актуальность разработки такого инструмента объясняется ростом объемов информации в сети Интернет, а также ростом аудитории сети.

В статье авторы проводят анализ проблемы поддержки принятия решений по противодействию нежелательной информации в сети Интернет. Описывают общий подход к противодействию нежелательной информации. Предлагают категорирование, схему сбора, предобработки и обработки данных для выделения нежелательной информации и последующего противодействия. Также в работе авторами вводится математическая модель системы поддержки принятия решений по противодействию.

Ключевые слова: нежелательная информация, информационный объект, защита от информации, поддержка принятия решений, контрмеры, противодействие

Введение

Информационная сфера включает все множество объектов и субъектов, осуществляющих сбор, формирование, распространение и использование информации [1]. На сегодняшний день она является неотъемлемой частью жизни общества и существенно влияет на его настроения и психическое здоровье нации. Свобода обмена информацией в сети Интернет создает новые возможности для взаимодействия пользователей, устройств и процессов. Все очевиднее становится проблема информационно-психологического ущерба от нежелательной информации. В связи с этим, одним из актуальных вопросов является защита от нежелательной информации, в том числе в сети Интернет [2]. В Российской Федерации (РФ) вопросы информационной безопасности регулируются Федеральным законом «Об информации, информационных технологиях и о защите информации» от 27.07.2006 N 149-ФЗ. Согласно пункту 6 статьи 10 данного закона в РФ запрещается распространение информации, которая направлена на пропаганду войны, разжигание национальной, расовой или религиозной ненависти и вражды, а также иной информации, за распространение которой предусмотрена уголовная или административная ответственность [3]. В настоящее время контроль и надзор за соблюдением этих требований ведется в основном вручную.

В то же время, Интернет активно развивается, и его аудитория постоянно растет. По статистике только доменов в зоне RU в РФ зарегистрировано 5 126 061, из них в июне 2018 года было зарегистрировано 108 471. В день в среднем появляется 3 616 новых доменов. В 44,2% случаев на домен прикреплен сайт, в 22,3% – одностраничный сайт [4]. Рост аудитории пользователей сети Интернет в РФ с 2004 по 2014 представлен на рисунке 1 [5].

Поэтому остро встает вопрос автоматизации поиска нежелательной информации, в том числе за счет разработки систем поддержки принятия решений по отнесению информации к нежелательной, и противодействия такой информации. Целью данной работы является исследование

[1] Младший научный сотрудник лаборатории проблем компьютерной безопасности.
[2] Кандидат технических наук, старший научный сотрудник лаборатории проблем компьютерной безопасности.

существующих решений по поиску и противодействию нежелательной информации и повышение их эффективности за счет разработки новых методов, моделей и методик.

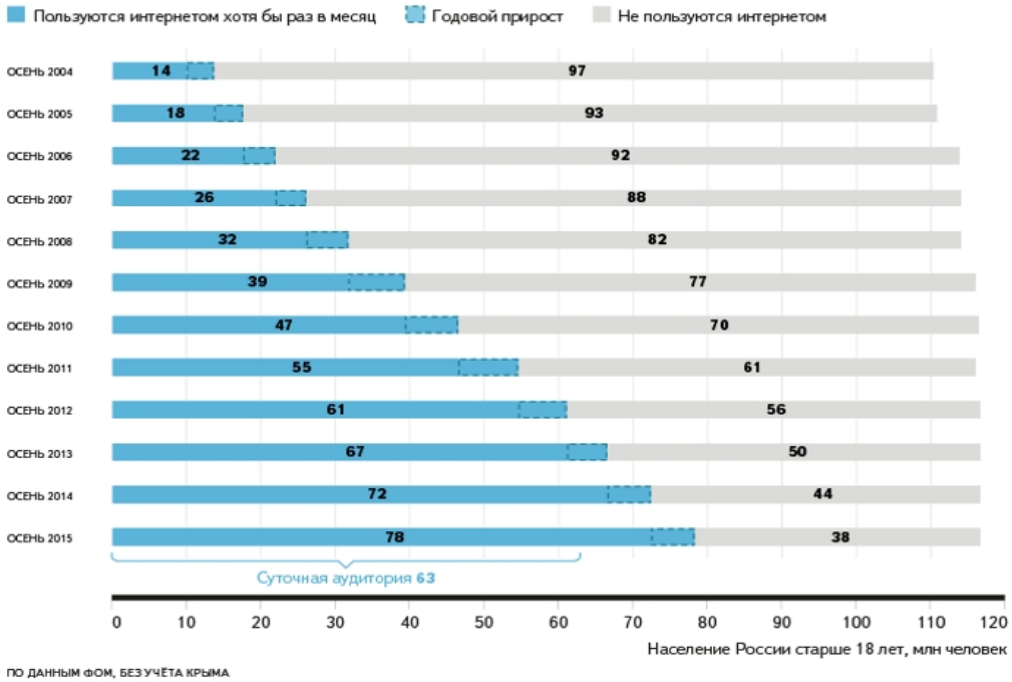


Рис. 1. Соотношение пользователей Интернета и людей, которые не выходят в сеть, в РФ

Анализ проблемы противодействия нежелательной информации

В настоящее время в РФ существует несколько механизмов защиты от нежелательной информации [6]. К ним относится: ручная классификация информационной продукции по категориям доступа к просмотру, при этом механизм защиты предполагает выдачу предупреждения о категории доступа; ручная блокировка ресурсов, содержащих нежелательную информацию; автоматизированные системы защиты от информации в виде отдельных опций в операционных системах (ОС).

Классификация информационной продукции осуществляется по категориям доступа к просмотру. В РФ выделяются следующие категории пользователей по возрасту: 1) от 0 до 6 лет; 2) от 6 до 12 лет; 3) от 12-16 лет; 4) от 18 лет. Согласно законодательству РФ, классификация информационной продукции осуществляется ее производителями и (или) распространителями самостоятельно. Защитные меры в данном случае носят предупредительный характер: производитель и (или) распространитель информации должен разместить на сайте знак категории, к которой отнесена информация; в случае, если информация относится к категории «от 18 лет», распространитель обязан предупредить пользователя и получить подтверждение, что он/она достиг соответствующего возраста [7].

Блокировка ресурсов направлена на защиту от запрещенного контента и осуществляется на основе реестра блокировок. Порядок внесения доменного указателя имени страницы и отнесения информации к категории, запрещенной определен Приказом от 18 мая 2017 года №84/292/351/ММВ-7-2/461С [8]. По сути, блокировка является способом сокрытия нежелательной информации от публичного просмотра. На сегодняшний день даже пользователи с невысоким уровнем технической грамотности могут обойти эту меру, достаточно воспользоваться услугами оператора связи не подконтрольного законодательству РФ.

Автоматизированные системы защиты от информации представляют собой средства фильтрации доменных указателей страниц, доступных пользователю. Требования к способам ограничения доступа к информационным ресурсам, а также требования к размещаемой информации определены приказом Роскомнадзора от 14.12.2017 г. [8]. Они доступны пользователям в виде отдельных опций в ряде ОС, таких как Windows, iOS (родительский контроль, защита от кон-

тента, содержащего ненормативную лексику и т.д.). Однако, в настоящий момент, они не распространены и не встроены в Linux, и что значительно опаснее, в ОС Android. По статистике в России большая часть пользователей для выхода в Интернет пользуется мобильными телефонами и планшетами. Так, по данным Mediascope, аудитория Интернет в период с октября 2016 по март 2017 года достигла 87 млн человек. Это более чем 70% от всего населения России. При этом с мобильных устройств выходили в Интернет хотя бы один раз в месяц 66 млн, что составляет 54% от населения страны. А 20 млн (16% от населения страны) пользуются сетью Интернет только через мобильные устройства. Быстрее всего растет аудитория на смартфонах. За указанный период аудитория выросла на 15% по отношению к предыдущему [9].

Данная статистика показывает, что огромная часть населения страны никак не защищена от воздействия нежелательной информации. Существующие меры в области защиты от нежелательной информации направлены на соблюдение нормативно-правовых актов. Нормативно-правовые акты не имеют публичных пояснений и комментариев по отнесению информации к нежелательной, а блокировка ресурсов является неэффективной с точки зрения ресурсозатрат и сокрытия нежелательной информации мерой. На сегодняшний день любая информация может быть доступна и эффективного инструмента защиты от нежелательной информации в Интернет нет. Поэтому существует необходимость разработки моделей, методов, методик и алгоритмов в области обеспечения поддержки принятия решений по отнесению информации к нежелательной и по противодействию такой информации.

Подход к противодействию нежелательной информации

Подход включает этап выделения нежелательной информации и этап противодействия нежелательной информации. На первом этапе предлагается использовать методику классификации информационных объектов.

Под информационным объектом будем понимать совокупность данных, трактуемых информационной системой как единое целое [10]. Все способы подачи информации, которыми может быть транслирована нежелательная информация, являются информационными объектами. Например, ресурс, пост, запись, сообщение, комментарий.

Для классификации информационных объектов будем использовать как свойства самого информационного объекта, так и следующие свойства: тип устройства доступа к информации, ОС пользователя, возрастная категория пользователя объекта. За основу классификации информационных объектов по типу устройства доступа к информации и возрастной категории пользователей объекта возьмем матрицу доступа пользователей, представленную в таблице 1.

Таблица 1

Матрица доступа пользователей к информации

№	Возраст пользователя	Тип устройства	
1	До 6 лет	Мобильный	Персональный компьютер (ПК)
2	От 6 до 12 лет	Мобильный	ПК
3	От 13 до 16 лет	Мобильный	ПК
4	От 17 до 18 лет	Мобильный	ПК
5	От 18 лет	Мобильный	ПК

Свойства информационного объекта опишем следующим образом. Пусть информационный объект обладает совокупностью качественных свойств. Любое i-е свойство информационного объекта может быть описано количественно с помощью некоторой переменной a_i, значение которой характеризует его качество относительно выбранного свойства. Эту переменную назовем показателем i-го свойства информационного объекта. Если ее можно представить в виде функциональной зависимости $a_i = f(a_1, a_2, ..., a_{i-1}, a_{i+1}, ..., a_n)$, то a_i называется обобщенным показателем свойств информационного объекта.

Определим следующие классы нежелательных информационных объектов (далее НИО): группа риска №1, группа риска №2. Группа риска №1 – это те НИО, на которые в приоритете должны быть направлены меры противодействия. НИО, входящие в группу риска №2, соответственно, являются менее приоритетными. В группу №1 попадают НИО, которые направлены на аудиторию мобильных устройств, социальных сетей. При настройке системы поддержки принятия решений (СППР) для выделения нежелательной информации, выбранные признаки классификации используются следующим образом. На первом этапе сегментируются ресурсы по способу доступа пользователей к ним. Например, социальные сети доступны пользователям через приложения, следовательно, данные информационные системы попадают в группу риска № 1. Если веб-страница не адаптирована под мобильные устройства, то ее можно отнести к группе риска № 2.

Для отнесения к группе риска № 1 или к группе риска № 2 авторы учитывают также свойства информационных объектов.

Для сбора и обработки информации, необходимой для классификации информационных объектов, авторами предлагается схема, представленная на рисунке 2. Предложенная схема позволит собирать сведения по следующим признакам: a_i – свойство информационного объекта; b_i – свойство пользователя; c_i – свойство среды передачи (устройство, ОС).

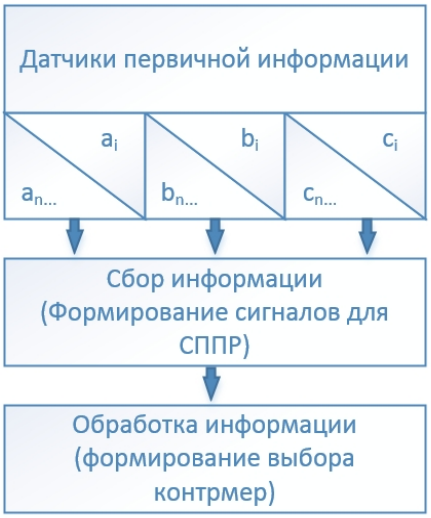


Рис. 2. Схема сбора и обработки информации в Интернет для поддержки принятия решений об отнесении информации к нежелательной

На втором этапе подхода предлагается использовать разработанную авторами систему противодействия нежелательной информации. Математическая модель предлагаемой системы противодействия нежелательной информации включает в себя:

- вектор (пространство) входных сигналов;
- вектор (пространство) состояний;
- совокупность входных сигналов;
- совокупность выходных сигналов;
- соотношения, связывающие входные сигналы, выходные сигналы и вектор состояний.

Величины, определяющие внешнее воздействие на систему поддержки принятия решений по противодействию нежелательной информации, называются ее входными сигналами. Входным сигналом для начала работы системы противодействия нежелательной информации может являться изменение свойств информационного объекта, в случае достижения свойствами определенных значений.

Положение или состояние системы противодействия нежелательной информации в каждый момент времени характеризуется состоянием ее подсистем, модулей сбора, их взаимным положением и воздействием. Эти характеристики называются параметрами системы, или перемен-

ными состояния системы. Под вектором или пространством состояний системы понимается вся совокупность переменных состояний (параметров) системы. Входные сигналы вызывают изменение состояния системы, направленное на получение выходных сигналов путем реализации методики поддержки принятия решений по противодействию ее подсистемами. Выходными сигналами называются величины, определяющие выбор контрмер.

Входные и выходные сигналы системы противодействия нежелательной информации, как некоторые функции времени и изменения вектора состояния во времени характеризуют функционирование системы, или ее поведение.

Для преобразования входных сигналов в выходные в рамках СППР используется методика противодействия нежелательной информации. Методика противодействия нежелательной информации должна учитывать свойства нежелательных информационных объектов, динамику их распространения, источник и аудиторию. Источником нежелательной информации является злоумышленник, осуществляющий мероприятия по ее распространению, в результате которых меняются свойства информационных объектов. Такие мероприятия могут быть представлены информационно-признаковой моделью в виде иерархического взвешенного графа типа «корневое дерево», что также позволит автоматизировать процесс противоборства с нежелательной информацией.

Будущие исследования будут посвящены выделению и группировке признаков воздействия со стороны злоумышленника, позволяющих сформировать модель мероприятий по распространению нежелательной информации и определить границы аудитории выбранной злоумышленником. В результате предполагается разработать адаптивную методику противодействия нежелательной информации, учитывающую вид информации, способ ее распространения и ее аудиторию. Для дальнейшей настройки и автоматизации системы поддержки принятия решений планируется применение нейросетевых технологий [11, 12].

Заключение

В работе проведен анализ проблемы противодействия нежелательной информации и разработан общий подход противодействия нежелательной информации. В рамках подхода предложена методика отнесения информации к нежелательной, разработана схемы сбора, предобработки и обработки данных. Также авторами описана математическая модель системы поддержки принятия решений по противодействию нежелательной информации и описаны направления будущих исследований в рамках разработки методики противодействия нежелательной информации.

Работа проводилась при поддержке гранта РНФ (проект № 18-11-00302) в СПИИРАН.

ЛИТЕРАТУРА

1. Доктрина информационной безопасности Российской Федерации (утв. Президентом РФ 09.09.2000 N Пр-1895).
2. Вильчинская Э.А., Виткова Л.А., Парсон И.М. Актуальные проблемы правового регулирования средств массовой информации //В сборнике: Право и информация: вопросы теории и практики сборник материалов международной научно-практической конференции. Сер. «Электронное законодательство» ФГБУ «Президентская библиотека имени Б. Н. Ельцина». 2017. С. 102-108.
3. Федеральный закон от 27.07.2006 N 149-ФЗ (ред. от 29.06.2018) «Об информации, информационных технологиях и о защите информации» (с изм. и доп., вступ. в силу с 30.06.2018) // «СЗ РФ», 31.07.2006, N 31 (1 ч.), ст. 3448.
4. Статистика доменов.RU и .РФ за июнь 2018. Режим электронного доступа: https://cctld.ru/ru/statistics/ (дата обращения 23.07.2018)
5. Виткова Л.А., Потехин И.Ю., Сахаров Д.В. Проблема выявления информационно-психологического воздействия в информационной инфраструктуре российской федерации // В сборнике: АПИНО 2017. Сборник научных статей VI Международной научно-технической и научно-методической конференции. В 4-х томах. Под редакцией С.В. Бачевского. 2017. Т. 1. С. 166-170.
6. Котенко И.В., Саенко И.Б., Чечулин А.А. Защита от нежелательной и вредоносной информации в глобальных информационных сетях // Информационно-психологическая и когнитивная безопасность: состояние и задачи. Коллективная монография / Под ред. И. Ф. Кефели, Р. М. Юсупова. СПб.: Изд-во «Аврора» 2017. С.207-229.
7. Федеральный закон «О защите детей от информации, причиняющей вред их здоровью и развитию» от 29.12.2010 N 436-ФЗ // «Собрание законодательства РФ», 03.01.2011, N 1, ст. 48.

8. Приказ Роскомнадзора от 14.12.2017 г. Режим электронного доступа: http://eais.rkn.gov.ru /docs/0001201706290015.pdf (дата обращения 23.07.2018).
9. Постановление правительства москвы от 14.06.2005 №439-пп «о дальнейшем проведении работ по созданию московского городского портала» (вместе с «концепцией системы городских порталов», «функциональными требованиями к московскому городскому порталу (мгп) в сети интернет»).
10. Аудитория пользователей интернета в России в 2017 году составила 87 млн. человек. Режим электронного доступа: http://mediascope.net/press/news/744498/ (дата обращения 23.07.2018).
11. Саенко И.Б., Котенко И.В., Скорик Ф.А. Мониторинг и прогнозирование состояния компьютерных сетей на основе применения гибридных нейронных сетей // Изв. вузов. Приборостроение, Т.59, № 10, 2016, С.795-800.
12. Браницкий А.А., Котенко И.В. Построение нейросетевой и иммуноклеточной системы обнаружения вторжений // Проблемы информационной безопасности. Компьютерные системы, № 4, 2015. С.23-27.

L.A. Vitkova, E.V. Doynikova (Saint-Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, Saint-Petersburg)

Decision support for protection against inappropriate information

Abstract. Currently the problem of information and psychological damage from inappropriate content is becoming obvious. Thus, protection against inappropriate information is significant. Modern measures in this area in Russian Federation usually consist in the development of regulatory measures and they are precautionary. Besides, modern regulations do not clarify attributing information to inappropriate information. Today there is no satisfactory tool for an automated determination and protection against inappropriate content in Internet. Increasing volume of information in Internet and of its audience shows the relevance of development of such tool.

In the study the authors analyze the problem of protection against inappropriate content in Internet. They describe a common approach to protection against inappropriate information. The paper proposes data categorization scheme and scheme of data gathering, preprocessing and processing. Besides, the authors introduce a mathematical model of decision support system and concept of the technique of protection against inappropriate content.

ПРИНЦИПЫ АДАПТАЦИИ И ПЕРЕОБУЧЕНИЯ СИСТЕМЫ АНАЛИЗА ИНФОРМАЦИОННЫХ ОБЪЕКТОВ ДЛЯ ЗАЩИТЫ ОТ НЕЖЕЛАТЕЛЬНОЙ И ВРЕДОНОСНОЙ ИНФОРМАЦИИ

В. А. Десницкий[1], **А. А. Браницкий**[2]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук,
С.-Петербург, 14-я линия Васильевского острова, 39,
desnitsky@comsec.spb.ru, branitskiy@comsec.spb.ru*)

Аннотация

Цель работы — разработка и анализ архитектуры системы классификации нежелательной и вредоносной информации по множеству классификационных признаков с использованием текстовой информации сайтов, html-данных, данных о структуре документов и URL-данных. В докладе представлено описание предлагаемого подхода для классификации html-страниц и архитектуры разработанной на его базе системы.

Ключевые слова: адаптация, переобучение, нежелательная и вредоносная информация, защита от информации

Введение

В настоящее время проблема классификации нежелательной и вредоносной информации требует эффективных решений как для повышения уровня защищенности конечных пользователей сети Интернет, в том числе в рамках систем родительского контроля, так и при организации механизмов выявления аномального трафика на уровне маршрутизаторов больших коммуникационных сетей. Ввиду практической ограниченности методов экспертного контроля нежелательной и вредоносной информации, связанной с необходимостью загрузки и анализа содержимого Web-сайтов вручную, все большую актуальность получают методы автоматизированной классификации Web-сайтов и их отдельных страниц с использованием методов машинного обучения. При этом эффективность применяемых для этих целей программно-технических решений должна определяться в первую очередь точностью выявления потенциально опасной информации, выражаемой в отсутствии или минимизации ошибок классификации первого и второго рода [1].

Цель работы заключается в разработке подхода и архитектуры системы анализа информационных объектов для защиты от нежелательной и вредоносной информации на примере решения задачи классификации Web-сайтов и отдельных размещенных на них html-страниц. Особенность предлагаемого подхода заключается в иерархическом построении решающего правила в виде набора базовых классификаторов и объединяющего их выходные данные верхнеуровневого классификатора. В работе представлено описание предлагаемого подхода, предназначенного для классификации содержимого html-страниц, а также архитектуры системы, разработанной на базе этого подхода.

Описание предлагаемого подхода

В работе предложен подход классификации нежелательной и вредоносной информации по N классификационным признакам с использованием текстовой информации сайтов, данных из определенного вида html-тегов Web-страниц, информации о структуре сайта, в том числе данных об их вложенности, а также URL-идентификаторов страниц, которые могут также успешно использоваться в процессе классификации. Нежелательность и вредоносность информации определяются принадлежностью соответствующих страниц Web-сайтов одной или нескольким определенным тематическим категориям.

Система классификации нежелательной и вредоносной информации включает ряд базовых классификаторов, каждый из которых позволяет на основе определенного типа используемых входных данных определить принадлежность конкретного образца одной из категорий в виде

[1]Кандидат технических наук, старший научный сотрудник лаборатории проблем компьютерной безопасности.
[2]Младший научный сотрудник лаборатории проблем компьютерной безопасности.

некоторого бинарного значения. Каждый такой классификатор проходит процесс обучения с выделением обучающих и контрольных выборок и вычислением показателей точности [2].

Полученные на основе применения бинарных классификаторов данные подаются в свою очередь в качестве входных данных аспектным классификаторам, которые в рамках определенного типа входных данных — текста, структуры тегов, URL-адресов и пр. — позволяют получить нормализованную вероятностную характеристику каждого образца исходных данных по серии категорий. После чего итоговый классификатор обобщает данные от аспектных классификаторов по всем имеющимся типам исходной информации и может использоваться в качестве основы предлагаемого программного средства классификации нежелательной и вредоносной информации.

Модульность предложенного подхода предполагает возможность гибкого добавления дополнительных базовых и аспектных классификаторов, которые позволили бы повысить точность итогового классификатора без необходимости полного переобучения всех используемых классификаторов. В частности, для ряда категорий точность распознавания сайтов может быть значительно повышена благодаря использованию базовых классификаторов изображений, содержащихся на страницах Web-сайтов. Таким образом, предложенный подход позволяет минимизировать временные затраты, связанные с изменением типов анализируемых исходных данных, модифицируя не систему классификации целиком, а лишь отдельное подмножество базовых классификаторов, соответствующий аспектный классификатор и итоговый классификатор, тем самым адаптируя имеющееся решение под изменившиеся требования.

Реализация предлагаемого подхода

В разработанной системе, предназначенной для классификации содержимого html-страниц, используется три уровня классификаторов [3]. Выходные значения классификаторов k-ого уровня передаются в качестве входных значений классификаторам (k+1)-ого уровня (k=1,2). Разработанный подход изображен на рис. 1.

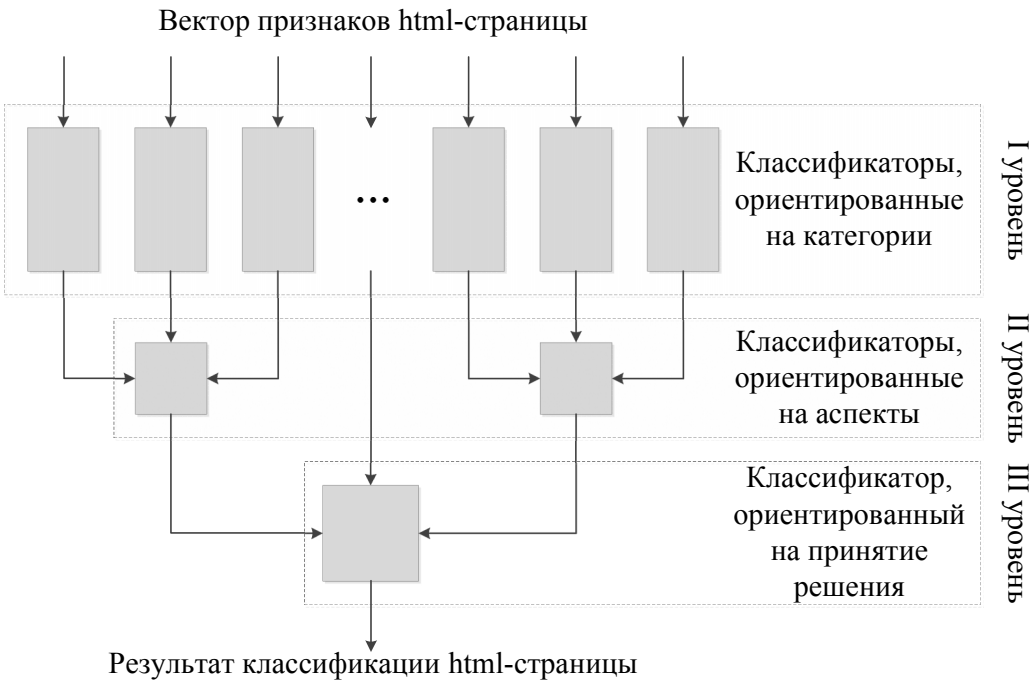


Рис. 1. Представление разработанного подхода в виде иерархии классификаторов

На первом уровне размещаются бинарные классификаторы, которые выполняют проверку принадлежности анализируемого объекта к определенной категории. При построении каждого

из таких классификаторов используются признаки, свойственные той или иной категории, что позволяет учитывать особенности, характерные для определенной группы исходных данных, и обучать классификаторы на том подмножестве атрибутов, которые наиболее полно отражают анализируемый контент в рамках рассматриваемой категории.

Второй уровень включает классификаторы, предназначенные для определения темы при использовании отдельных структурных элементов (аспектов) html-страницы, к примеру URL-адреса или элементов форматирования html-страницы [4]. Обрабатываемый классификаторами этого уровня вектор формируется на основе выходных значений, полученных от бинарных классификаторов первого уровня.

На третьем уровне располагается классификатор, который выполняет роль итоговой классификации анализируемой html-страницы. Для повышения эффективности функционирования этого классификатора его входными значениями являются не только выходные данные классификаторов второго уровня, но и исходные признаки анализируемых html-страниц.

Преимущества описанного иерархического подхода классификации содержимого html-страниц заключаются в следующем:

- Появляется возможность многократного и параллельного анализа структуры и содержимого html-страницы за счет декомпозиции системы в виде множества классификаторов, расположенных на нескольких уровнях, что в конечном итоге позволяет повысить точность классификации;
- Снижаются объемы затраченных системных и временных ресурсов за счет наличия в системе легковесных бинарных классификаторов первого уровня;
- Появляется возможность гибкой модификации набора классификаторов, включая их изменение, добавление или удаление, что, в частности, позволяет выполнять процедуры переобучения и адаптации отдельных классификаторов.

Архитектура разработанной системы

Прототип реализован с использованием языка программирования Java [5], что предоставляет возможность кроссплатформенной переносимости кода между различными средами (включая различные ОС и архитектуры компьютеров), обеспечивает высокий уровень безопасности исполняемых инструкций за счет предоставления высокоуровневых операторов и генерации управляемого при помощи JVM кода. В качестве инструмента для построения и обучения классификаторов был использован RapidMiner [6]. Выбор данного программного продукта обоснован удобством и быстротой создания готовых к обучению моделей классификаторов, а также наличием API для его интеграции с Java-приложениями. Архитектура разработанной системы классификации нежелательной и вредоносной информации, содержащейся на html-страницах, представлена на рис. 2.



Рис. 2. Архитектура разработанной системы классификации содержимого html-страниц

Сперва исходные наборы данных, представленные в виде файлов html-страниц, обрабатываются компонентом извлечения признаков. Формируемые этим компонентов наборы признаков помещаются в хранилище данных, которое представляет собой SQL-ориентированную базу данных. Компонент анализа данных представляет собой отдельное Java-приложение, построенное на основе вышеупомянутого инструментального средства RapidMiner. На последнем этапе выполняется генерация отчетов, содержащих такие показатели эффективности классификации html-страниц, как точность, полноту и F-меру.

Заключение

В докладе был описан подход к анализу информационных объектов для защиты от вредоносной и нежелательной информации на примере классификации html-страниц. Особенность рассмотренного подхода заключается в иерархическом размещении классификаторов, выполняющих различные функции для определения принадлежности анализируемой html-страницы к той или иной категории. Архитектура разработанной системы построена таким образом, чтобы стал возможным процесс переобучения классификаторов с использованием новых данных и их адаптации с учетом изменившихся данных.

Направление дальнейших исследований — совершенствование предложенной архитектуры системы классификации нежелательной и вредоносной информации и проведение экспериментов, направленных на вычисление показателей эффективности классификации нежелательной и вредоносной информации, представленной на html-страницах, и измерение объемов потребляемых системных и временных ресурсов.

Работа выполнена при финансовой поддержке гранта РНФ (проект № 18-11-00302) в СПИИРАН.

ЛИТЕРАТУРА

8. Desnitsky V., Kotenko I. Expert knowledge based design and verification of secure systems with embedded devices // Lecture Notes in Computer Science. 2014. Vol. 8708. P. 194–210.
9. Desnitsky V.A., Kotenko I.V., Nogin S.B. Detection of anomalies in data for monitoring of security components in the internet of things // Proceedings of International Conference on Soft Computing and Measurements, SCM 2015. 2015. P. 189–192.
10. Kotenko I., Chechulin A., Komashinsky D. Categorisation of web pages for protection against inappropriate content in the internet // International Journal of Internet Protocol Technology (IJIPT). 2017. Vol. 10, no. 1. P. 61–71.
11. Котенко И.В., Чечулин А.А., Комашинский Д.В. Автоматизированное категорирование веб-сайтов для блокировки веб-страниц с неприемлемым содержимым // Проблемы информационной безопасности. Компьютерные системы. 2015. № 2. С. 62–68.
12. Java. Электронный ресурс. http://www.oracle.com/technetwork/java/javase/overview/index.html.
13. RapidMiner. Электронный ресурс. https://rapidminer.com/.

V.A. Desnitsky, A.A. Branitskiy (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg)

Principles of adaptation and retraining of a system for analyzing information objects for protection against inappropriate and malicious information

Abstract. The purpose of the paper is the development and analysis of the architecture of the classification system of inappropriate and malicious information on a variety of classification characteristics using textual information of sites, html-data, data on the structure of documents and URL-data. The report describes the proposed approach for classification of html-pages and the architecture of the system developed on its basis.

ПОДХОД К ПРОГНОЗИРОВАНИЮ И РЕАГИРОВАНИЮ НА КИБЕРАТАКИ В ИНДУСТРИАЛЬНОМ ИНТЕРНЕТЕ ВЕЩЕЙ НА ОСНОВЕ НЕЙРОНЕЧЁТКИХ СЕТЕЙ

Е. В. Дойникова[1]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), 199178, Россия, С.-Петербург, 14-я линия В.О., д. 39, (812) 328-71-81, doynikova@comsec.spb.ru*)

Аннотация

Актуальность разработки новых подходов к прогнозированию и реагированию на кибератаки в индустриальном Интернете вещей объясняется, с одной стороны, его активным развитием, а с другой стороны, отсутствием удовлетворительных решений по обеспечению информационной безопасности в области таких систем. В работе приводится общее описание предлагаемого подхода, основанного на применении нейронечетких сетей, и его основных элементов. Наибольшее внимание уделяется входным данным для обучения нейронечеткой сети. Для этого предлагается применение комплекса показателей, представленного в виде онтологии.

Ключевые слова: индустриальный Интернет вещей, кибератаки, прогнозирование, поддержка принятия решений, онтология метрик.

Введение

Индустриальный (или промышленный) Интернет вещей, с одной стороны, расширяет возможности пользователей, делая сервисы, информацию и продукты доступнее через связи между различными подключенными «умными» устройствами и процессами, а с другой стороны, создает новые возможности для кибератак. Разнообразие протоколов, устройств, архитектур в рамках систем индустриального Интернета вещей усложняет его защиту.

Анализ предметной области и релевантных работ показал, что в настоящее время область защиты систем индустриального Интернета вещей только начинают развиваться [1]. Учет требований безопасности при разработке элементов индустриального Интернета вещей, разработка и повсеместное внедрение безопасных протоколов и программного и аппаратного обеспечения займет достаточно длительное время. Поэтому, на текущий момент одной из важнейших задач обеспечения безопасности промышленного Интернета является мониторинг безопасности, подразумевающий сбор и обработку данных безопасности, прогнозирование развития кибератак и принятие решений по реагированию на основе собранных данных.

Ввиду больших объемов данных безопасности, разнородности элементов рассматриваемых систем и связей между ними, а также их открытости, системы мониторинга безопасности для индустриального Интернета вещей нуждаются в интеллектуализации и ряде усовершенствований. Поэтому целью данной работы является повышение защищенности систем индустриального Интернета вещей за счет разработки подхода к прогнозированию развития кибератак и реагированию на кибератаки в индустриальном Интернете вещей. Для достижения поставленной цели в работе предлагается подход к прогнозированию и реагированию на кибератаки в индустриальном Интернете вещей на основе нейронечетких сетей.

Общее описание подхода

Основные требования к разрабатываемому подходу к прогнозированию и реагированию на кибератаки включают применимость для систем индустриального Интернета вещей, что объясняется анализируемой предметной областью; адаптивность, что подразумевает применимость в условиях изменчивой архитектуры системы; возможность обработки разнородных данных безопасности, что объясняется разнородностью элементов индустриального Интернета вещей; формирование на основе доступных данных значимых показателей, применимых для решения задач прогнозирования развития атак и принятия решений; возможность учета новой информации для

[1] Кандидат технических наук, старший научный сотрудник лаборатории проблем компьютерной безопасности.

усовершенствования механизмов прогнозирования и принятия решений; оперативность, для своевременного реагирования на кибератаки. Перечисленные требования были сформулированы с учетом основных особенностей систем индустриального Интернета вещей, в том числе, больших объемов данных безопасности, разнородности элементов рассматриваемых систем и связей между ними, и их открытости. Для достижения поставленной цели с учетом введенных требований в работе предлагается использовать методы интеллектуального анализа данных, а именно, нейронечеткие сети для прогнозирования целей кибератак и генетические алгоритмы для выбора контрмер. Подход включает следующие основные этапы: этап определения возможных классов целей кибератак, этап определения целей атак, и этап реагирования на кибератаки. В данной работе рассматривается только первый этап определения возможных классов целей кибератак. Для определения возможных классов целей кибератак, выраженных нечеткими множествами, предлагается применение онтологии метрик защищенности и механизма нечеткой логики. В дальнейшем, для настройки параметров функций принадлежности нечеткому множеству, предлагается использовать вычислительные возможности нейронных сетей и их способность к обучению.

Онтология метрик защищенности

Онтология метрик защищенности позволяет определить связь метрик защищенности с различными классами целей атак. Предлагаемая онтология частично основана на онтологии метрик защищенности [2] и онтологии данных безопасности [3], предложенных авторами ранее. Для ее определения необходимо выделить множество концептов, то есть набор объектов, участвующих в процессе прогнозирования целей кибератак, их признаков, возможных значений признаков, а также связи между объектами с учетом их взаимоотношений. Для определения множества концептов разрабатываемой онтологии, выделим объекты прогнозирования целей кибератак с учетом следующих групп: основные объекты, участвующие в процессе кибератаки; источники входных данных; параметры объектов или метрики.

В кибератаке участвуют следующие основные объекты: источник угрозы, угроза, уязвимость, актив, владелец актива, воздействие, защитная мера, и риск. Ранее, в [2], авторами выделялись объекты с точки зрения основных взаимодействующих сущностей процесса оценки защищенности: атакующий (как источник угрозы), атака (как реализация угрозы), уязвимость, система, хосты и приложения (как активы), защитная мера, риск (в виде метрики), ущерб (в виде метрики, определяющий воздействие), и предупреждение (для учета изменения ситуации по защищенности в динамике). В [3] объекты выделялись для создания хранилища данных безопасности. Были определены следующие концепты онтологии: конфигурация, продукт, эксплоит, слабое место, атака, уязвимость, контрмера. Кроме этих двух онтологий, в данной работе использовалась развернутая таксономия, предложенная в [4]. Она связывает в рамках инцидента безопасности атакующих с целями атак, и определяет классы выделенных сущностей. В рамках предложенной таксономии авторы выделяют «инцидент», который характеризуется такими объектами как «атакующий», «атака» и «цель». В свою очередь, атака характеризуется такими объектами как «инструмент атаки», «уязвимость», «событие» и «несанкционированный результат». Событие определяется сущностями «действие» и «целевой объект». Для каждого объекта нижнего уровня («атакующий», «инструмент атаки», «уязвимость», «действие», «целевой объект», «несанкционированный результат», «цель») авторы выделяют подклассы.

Онтология, разработанная в данном исследовании для выделения целей кибератак, включает следующие основные сущности: система, целевой объект, атакующий, слабое место, атака, уязвимость, контрмера, инцидент, цель атаки, инструмент атаки, событие, атакующее действие (рисунок 1). В данном исследовании рассматриваются системы индустриального Интернета вещей. При этом целевым объектом атаки может быть программно-аппаратное обеспечение таких систем, каналы связи между ними и соответствующие протоколы передачи данных, а также данные. Перечисленные концепты, определяющие основные объекты, участвующие в кибератаке, объединяются на рисунке 1 корневой сущностью Thing.

Для последующего определения цели атаки в онтологию добавлены подклассы объектов, предложенные в [4]. Часть подклассов представлена на рисунке 1 (закрашенные овалы), однако некоторые классы опущены, чтобы не перегружать рисунок.

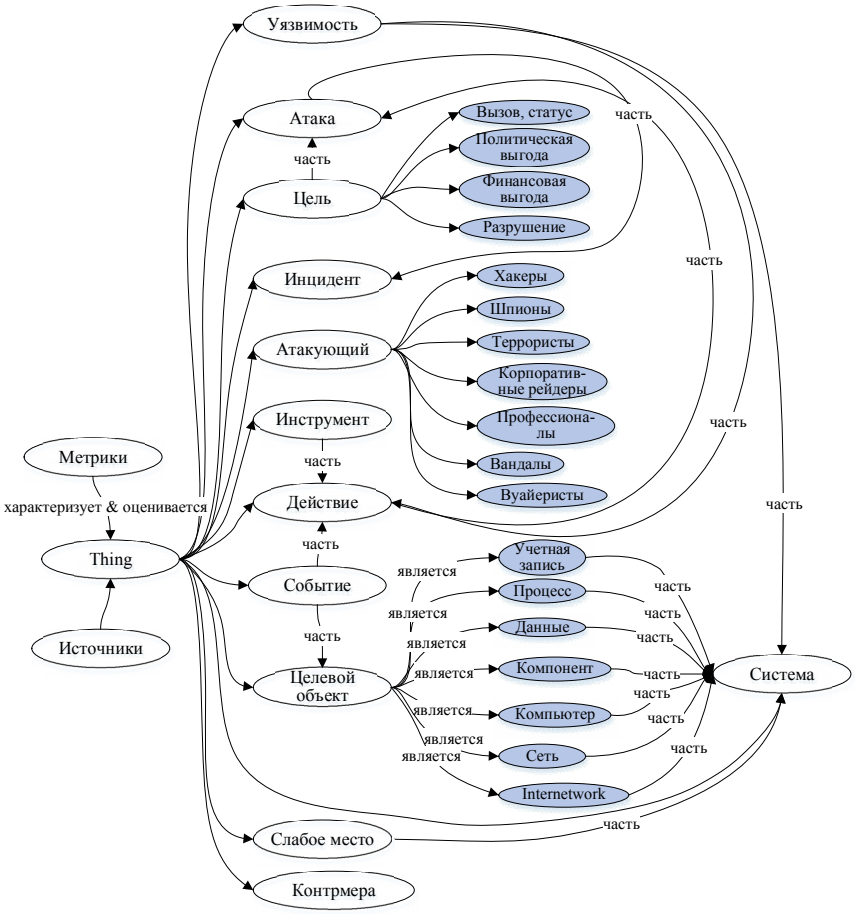


Рис. 1. Фрагмент онтологии определения целей кибератак, включающий концепты верхнего уровня

Группа концептов, определяющих источники входных данных, описана в [3]. Она объединяет открытые базы данных безопасности, такие как база слабых мест CWE, база шаблонов атак CAPEC, базы уязвимостей CVE и NVD, база эксплоитов ExploitDB, и другие. В рамках разрабатываемой онтологии они объединяются корневой сущностью «Источники», связанной с сущностью Thing, так как позволяют определить конкретные данные концептов Thing. Кроме того, концепты группы «Источник» используются как входные данные для другой группы объектов онтологии, определяющей параметры объектов. Для связи параметров (признаков) и объектов отдельно выделим концепт «Метрики». Введение данного концепта позволит определить класс, к которому относится цель атаки, путем выделения признаков и возможных значений признаков, присущих каждому классу. Ранее автором выделялись следующие классы метрик: системы, атак, атакующего, событий, выбора контрмер и интегральные [7]. Взаимосвязи между классами метрик и концептами онтологии, соответствующими объектам, представлены на рисунке 2 (связи вида «часть» не отображены для улучшения зрительного восприятия). Каждый класс метрик включает набор показателей, позволяющих определить класс характеризуемого ими объекта. Метрики выбираются исходя из целей оценки защищенности, параметров объекта, и его связей с другими объектами. Так, например, для атакующего можно выделить следующие параметры: «Уровень навыков атакующего» и «Знания атакующего о системе». «Уровень навыков атакующего» является интегральным показателем, рассчитываемые с использованием показателей других объектов. В том числе, «Сложность инструмента», «Сложность шагов атаки», «Успешность шагов атаки», «Степень сокрытия следов». Все эти показатели определяются на основе связи между атакующим и проводимыми им шагами атак Атака в свою очередь характеризуется показателями: «Вероятность атаки», «Ущерб». «Вероятность атаки» определяется сложностью выполнения и глубиной атаки, которые, в свою очередь, определяются показателями шагов, входящих в анализируемую многошаговую атаку.

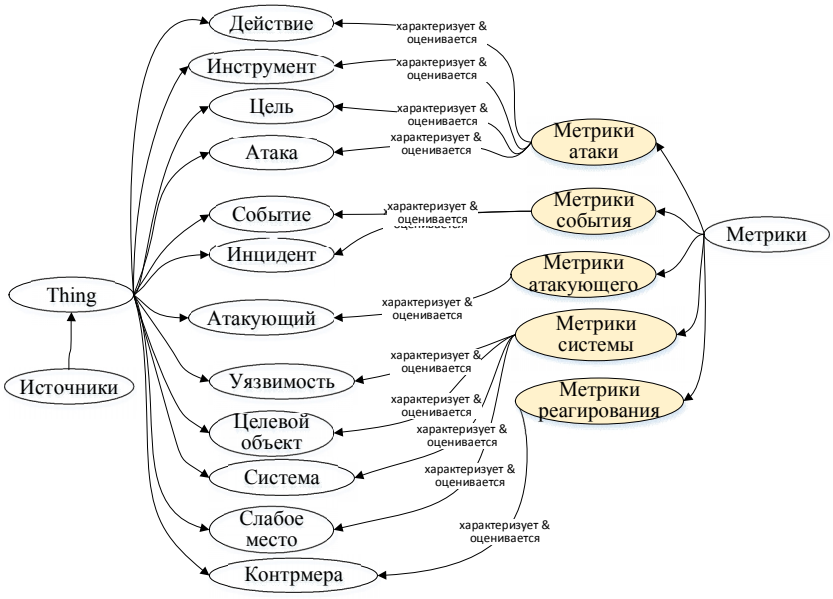


Рис. 2. Фрагмент онтологии, включающий классы метрик верхнего уровня

Взаимоотношения между объектами на рисунке 1 и рисунке 2 определяются видом связи. Классы и подклассы связаны стрелками без подписи. Например, класс «атакующий» и его подклассы «хакеры», «шпионы», «террористы», и другие. Также выделяются отношения вида «часть-целое» (связь «часть» на рисунке 1). Таким видом отношений связаны «Система» и ее компоненты, такие как «учетная запись», «процесс», «данные», «компонент», «компьютер», «сеть», «internetwork», «конфигурация», и «продукт». Взаимоотношения между метриками и объектами определяются следующими типами связей [2]: для отображения того, что метрика характеризует объект, используется связь «характеризует»; для определения того, что метрика вычисляется с использованием объекта, применяется связь «оценивается»; связь «характеризует & оценивается» обозначает двустороннюю связь; для определения метрик, вычисляемых на основе других метрик, применяется связь «функция».

Пример применения

В первом грубом приближении можно определить следующее соответствие между выделенными на рисунке 1 целями атак и качественными признаками атак, атакующего и целевого объекта. Для цели «политическая выгода», типы целевого объекта атаки включают сервисы, компьютеры и веб-сайты государственных организаций и отдельных политиков, финансовые центры, банки, и другие. Она будет характеризоваться серьезными нарушениями целостности информации, доступности информации, или конфиденциальности информации. Исполнителями таких атак (атакующими) чаще всего выступают организованные «профессионалы», так как для данных атак характерна высокая стоимость и сложность применяемых инструментов, сложность реализуемых действий, организованность шагов, успешность действий, а также грамотное заметание следов. Кроме того, они зачастую берут на себя ответственность за произведенную атаку. В отличие от цели «политическая выгода», цель «вызов, статус» может отличаться неорганизованностью, может быть привлекательна для разного рода злоумышленников от профессионалов до новичков, она будет характеризоваться разрозненностью шагов, и попытками нанесения максимального ущерба. Цель, связанная с получением финансовой выгоды, скорее привлечет профессионалов, хорошо заметающих следы, выполняющих самое простое действие из возможных, но способных выполнять сложные действия, высокими затратами, а также попытками получения максимальных привилегий и скрытия следов в системе до момента достижения цели. Для цели «разрушение» будет характерно нанесение максимальных разрушений на каждом шаге.

Для дальнейшего выделения целей на основе признаков в данном исследовании предполагается использовать реализацию нейронечеткой сети на основе ANFIS [5, 6] ввиду ее эффектив-

ности и оперативности при получении выводов. Эффективность основывается на том факте, что нейронечеткие сети являются универсальными аппроксиматорами [6]. Оперативность достигается за счет гибридного правила обучения, предложенного в [5]. Данная модель включает пять слоев, из которых настройке подлежит первый слой (настраиваются параметры функций принадлежности по методу градиентного спуска) и четвертый слой (настраиваются коэффициенты многомерной линейной регрессии по методу наименьших квадратов). Применение нейронечеткой сети для реализации предложенного подхода, а также формирование обучающей выборки для определения целей кибератак будет подробно рассмотрено в будущих исследованиях.

Заключение

В работе описана общая идея подхода к определению целей кибератак на основе онтологии метрик защищенности и нейронечетких сетей. Представлена первичная онтология определения целей и соответствующих метрик безопасности. В будущем планируется расширить предлагаемую онтологию с точки зрения применяемых метрик, определить их возможные значения, а также определить механизм применения нейронечеткой сети для выделения различных классов целей кибератак.

Работа проводилась при поддержке стипендии президента РФ (СП-751.2018.5).

ЛИТЕРАТУРА

1. Дойникова Е.В. Методики и средства реагирования на кибератаки в системах индустриального Интернета вещей // VII Международная научно-техническая и научно-методическая конференция "Актуальные проблемы инфотелекоммуникаций в науке и образовании" (АПИНО-2018). 28 февраля - 01 марта 2018 г. Сборник научных статей. Том 1. СПб., 2018.
2. Kotenko I., Saenko I., Polubelova O., Doynikova E. The ontology of metrics for security evaluation and decision support in SIEM systems. // In Proceedings of the 2nd International Workshop on Recent Advances in Security Information and Event Management, in conjunction with the 8th International Conference on Availability, Reliability and Security, Regensburg, Germany. IEEE Computer Society, 2013, pp. 638-645.
3. Kotenko I., Doynikova E., Fedorchenko A., Chechulin A. An ontology-based hybrid storage of security information. // In Information Technology and Control, 2018, in press.
4. Howard J., Longstaff, T. A common language for computer security incidents. Sandia Intelligence Labs, 1998.
5. Браницкий А.А. Иерархическая гибридизация бинарных классификаторов для выявления аномальных сетевых соединений // Труды СПИИРАН. 2017. Вып. 52. С. 204-233.
6. Jang J.S. ANFIS: adaptive-network-based fuzzy inference system // IEEE transactions on systems, man, and cybernetics. 1993. Vol. 23, no. 3. P. 665-685.
7. Kotenko I., Doynikova E. Countermeasure selection in SIEM systems based on the integrated complex of security metrics. // In Proceedings of the 23rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP 2015), 2015.

E.V. Doynikova (Saint-Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, Saint-Petersburg)

Approach to the cyberattack prediction and response based on the neuro-fuzzy networks in the industrial Internet of Things

Abstract. The relevance of development of novel approaches to cyberattack prediction and response in the industrial Internet of Things follows, on the one hand, from its active development, and on the other hand, from an absence of the satisfied solutions in the area of security of such systems. The paper provides common description of the proposed approach based on the neuro-fuzzy networks and of its main components. Main attention is paid to the input data for neuro-fuzzy network training. For this goal the study suggests application of the metrics set represented as an ontology.

ПРИМЕНЕНИЕ ОНТОЛОГИЧЕСКОГО ХРАНИЛИЩА ДЛЯ АНАЛИЗА ЗАЩИЩЕННОСТИ ИНФОРМАЦИОННЫХ СИСТЕМ

Е. В. Дойникова, **А. В. Федорченко**
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), 199178, Россия, С.-Петербург, 14-я линия В.О., д. 39, (812) 328-71-81, doynikova@comsec.spb.ru, fedorchenko@comsec.spb.ru*)

Аннотация

Одной из актуальных задач в настоящее время является обеспечение безопасности информационных систем. Ее решение невозможно без тщательного анализа защищенности системы. В работе предлагается использовать возможности онтологического хранилища данных безопасности для определения списка возможных угроз безопасности. Разработана онтологическая модель, объединяющая основные компоненты безопасности. Для выявления и прогнозирования возможных угроз и мер противодействия определена методика вывода новых знаний на основе связей онтологической модели. Проанализированы свойства уязвимостей и слабых мест информационных систем и их связь с угрозами безопасности. На основе проведенного анализа сформирован пример, демонстрирующий работу предложенного подхода, а именно, получение новых знаний о возможных угрозах.

Ключевые слова: анализ защищенности, онтология, онтологическое хранилище данных безопасности, угрозы безопасности, показатели защищенности.

Введение

Одной из актуальных задач в настоящее время является обеспечение безопасности информационных систем. Ее решение невозможно без тщательного анализа защищенности системы. Несмотря на многочисленные усилия в данной области, остается много нерешенных проблем. В том числе это относится к задачам выявления и прогнозирования возможных угроз. Целью данной работы является повышение защищенности информационных систем за счет применения онтологического хранилища данных безопасности при анализе защищенности.

В работе предлагается использовать возможности онтологического хранилища данных безопасности для определения списка возможных угроз безопасности. Преимуществом онтологического подхода, лежащего в основе разработанного хранилища, является отображение взаимосвязей между различными компонентами безопасности с использованием средств дескрипционной логики. Это позволяет получать ответы на различные вопросы безопасности путем онтологического вывода.

Для разработки онтологического хранилища данных безопасности в работе проанализированы открытые базы уязвимостей, слабых мест, угроз безопасности информационных систем, и другие, а также связи между соответствующими данными безопасности. Анализ проводился с использованием открытых стандартов представления данных безопасности. Предполагается, что разрабатываемое онтологическое хранилище данных безопасности позволит не только извлекать данные безопасности, но и получать новые знания об уровне защищенности системы на основе связей онтологической модели, лежащей в его основе. В работе предложена методика извлечения новых знаний безопасности с использованием онтологического хранилища данных безопасности и приведен пример ее применения для выявления возможных угроз безопасности и выбора защитных мер.

Онтологическое хранилище данных безопасности и методика извлечения новых знаний безопасности

Основным недостатком существующих баз данных безопасности является невозможность формирования комплексной картины безопасности из-за их разрозненности. Онтологический подход был взят за основу разрабатываемой базы данных безопасности, так как позволяет отображать сложные взаимосвязи между данными средствами дескрипционной логики. Разработанная в результате онтологическая модель данных безопасности включает набор концептов

анализируемой предметной области и взаимосвязей между ними. Концепты разработанной онтологической модели были определены с учетом следующей информации безопасности: уязвимости, программное и аппаратное обеспечение, слабые места программного и аппаратного обеспечения, эксплойты, шаблоны атак, конфигурации программного и аппаратного обеспечения, и контрмеры. Для определения взаимосвязей между этими объектами были проанализированы основные открытые базы данных безопасности и связи между ними. На Рис. 1 представлена иерархия наследования классов онтологической модели для гибридного хранилища информации безопасности. В качестве корня иерархии (сущность «Thing» в соответствии со стандартом OWL2 [1]) представлена информация безопасности.



Рис. 1. Иерархия наследования классов онтологической модели гибридного хранилища информации безопасности

Все классы разработанной онтологической модели условно разделяются на:

(1) типы информации безопасности (все кроме классов «Source» и «Reference» и их наследников):
   - контрмеры («Countermeasure» на Рис. 1);
   - уязвимости («Vulnerability» на Рис. 1);
   - атаки («Attack» на Рис. 1);
   - слабые места («Weakness» на Рис. 1);
   - эксплойты («Exploit» на Рис. 1);
   - программное и аппаратное обеспечение («Product» на Рис. 1);
   - конфигурации («Configuration» на Рис. 1);

(2) источники информации безопасности («Source» и дочерние классы):
- базы контрмер («CRE» [2] на Рис. 1);
- базы уязвимостей («X-Force» [3], «OSVDB» [4], «CVE» [5], и «NVD» [6] на Рис. 1);
- базы шаблонов атак («CAPEC» [7] на Рис. 1);
- базы слабых мест («CWE» [8] на Рис. 1);
- базы эксплойтов («Exploit-DB» [9] на Рис. 1);
- базы программного и аппаратного обеспечения («CPE» [10] на Рис. 1);
- базы конфигураций («CCE» [11] на Рис. 1);

(2) ссылки на сторонние источники описания конкретной информации безопасности («Reference»).

Данная модель описывает не только родительские отношения между понятиями области знаний «информация безопасности», но также учитывает отношения принадлежности между концептами (типами) и конкретными источниками данных. Однако иерархия наследования классов не отражает связи между сущностями (свойства-объекты) и возможных вариантов описания их индивидов (свойства-значения). Рассмотрим определение объектных свойств на примере уязвимостей. Выделено 4 иррациональных свойства: (1) «реализуетсяПосредством»; (2) «реализуетсяВ», (3) «реализуетсяПри», (4) «реализует». На основе указанных объектных свойств справедливы следующие аксиомы для сущности «Уязвимость»: (1) «реализуетсяПосредством эксплойта»; (2) «реализуетсяВ платформе»; (3) «реализуетсяПри уязвимойКонфигурации»; (4) «реализует слабость»; (5) «реализует атаку». Перечисленные свойства характеризуют отношения уязвимостей к пяти другими видам информации.

Также существуют отношения между экземплярами классов информации безопасности, не связывающие концепты между собой напрямую. Данное объектное свойство характеризует связь субъекта с внешним источником (например, по ссылке на URL или иному условно-уникальному идентификатору), выраженным сущностью «Reference».

Часть сущностей, характеризующих индивидов в онтологической модели, реализовано за счет определения свойств данных. Подобными свойствами являются поля описания информации безопасности. Например, для описания уязвимостей используются метрики CVSS [12], для которых задаются диапазоны возможных значений.

На основе разработанной модели предлагается методика извлечения новых знаний для оценки защищенности и выбора контрмер, включающая следующие шаги:

(3) сбор доступных данных безопасности;

(4) расширение данных безопасности путем обхода сущностей на основе связей между концептами онтологии;

(5) анализ данных (включая анализ значений свойств данных концептов).

. Для выявления и прогнозирования возможных угроз определен алгоритм вывода на основе связей онтологической модели. В работе также описывается каким образом предложенная модель используется при формировании онтологического хранилища данных безопасности. Проведены эксперименты, демонстрирующие работу предложенного подхода, а именно, получение новых знаний о возможных угрозах.

Применение данного алгоритма для выявления угроз безопасности и выбора защитных мер демонстрируется в следующем разделе.

Применение разработанной методики для выявления угроз безопасности и выбора защитных мер

Рассмотрим применение разработанной онтологии и хранилища данных безопасности для выявления угроз безопасности и выбора защитных мер (Рис. 2). В примере используются отношения между базой программного и аппаратного обеспечения (NVD), базой уязвимостей (NVD), базой слабых мест (CWE) и базой шаблонов атак (CAPEC).

На первом шаге система мониторинга безопасности сообщает об инциденте безопасности. Информация об инциденте безопасности включает информацию о скомпрометированной рабо-

чей станции и виде ущерба (это может быть получение привилегий нелегитимным пользователем и/или нарушение конфиденциальности / целостности / доступности информации). На основе этой информации можно выделить уязвимости рабочей станции в формате CVE, которые ведут к указанным последствиям.

В анализируемом примере (Рис.2) был выявлен нелегитимный доступ к веб-серверу и частичное нарушение конфиденциальности / целостности / доступности информации на нем.

На основе анализа программного и аппаратного обеспечения веб-сервера было обнаружено, что к таким последствиям ведет эксплуатация уязвимости CVE-2009-0689 [5] в программном обеспечении FreeBSD 7.2 (запись в формате CPE: cpe:2.3:o:freebsd:freebsd:7.2:*:*:*:*:*:*:* [10]).

В базе NVD [6] данная уязвимость связана со слабым местом CWE-119 в базе CWE. В свою очередь, это слабое место связано со следующими шаблонами атак: CAPEC-10 «Переполнение буфера через переменные среды», CAPEC-100 «Переполнение буфера», CAPEC-14 «Переполнение буфера инициализированное инъекцией на стороне клиента», CAPEC-24 «Отказ фильтра из-за переполнения буфера», и др. Все связанные шаблоны являются наследниками шаблона CAPEC-100 «Переполнение буфера».

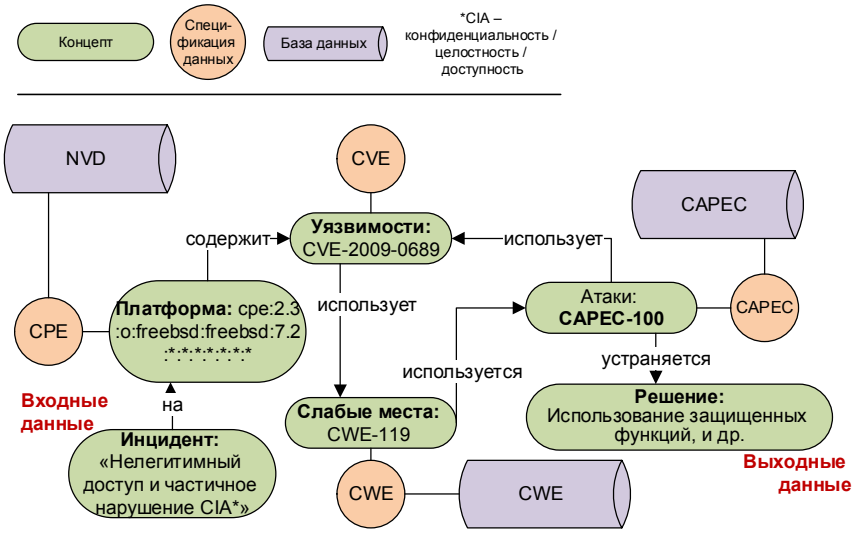


Рис. 2. Пример использования онтологической модели

Чтобы избежать атак этого типа в будущем в базе данных CAPEC предлагаются следующие решения (поле «Решения»): использовать язык или компилятор, осуществляющий автоматическую проверку границ; использовать функции, неуязвимые к переполнению буфера; при использовании небезопасных функций убедиться, что выполняется проверка границ; и др.

Необходимо отметить, что разработанный подход не предназначен для устранения определенной уязвимости или реагирования на конкретную атаку, но позволяет избежать обнаруженную проблему в будущем.

Заключение

В работе предложена концепция онтологического хранилища данных безопасности для определения списка возможных угроз безопасности. Разработана онтологическая модель данных безопасности. Для выявления возможных угроз и выбора мер защиты определена методика вывода новых знаний на основе связей онтологической модели. Проанализированы свойства уязвимостей и слабых мест информационных систем и их связь с угрозами безопасности. На основе проведенного анализа разработан пример, демонстрирующий работу предложенного подхода, а именно, получение новых знаний о возможных угрозах и противодействия им.

Работа выполнена при поддержке гранта президента РФ (МК-314.2017.9).

ЛИТЕРАТУРА

1. Официальный веб-сайт W3C. Web ontology language overview. Режим электронного доступа: https://www.w3.org/TR/owl-features/ (дата обращения 23.07.2018).
2. Официальный веб-сайт Common Remediation Enumeration (CRE). Режим электронного доступа: https://scap.nist.gov/specifications/cre/ (дата обращения 23.07.2018).
3. Официальный веб-сайт проекта IBM X-Force. Режим электронного доступа: http://xforce.iss.net (дата обращения 23.07.2018).
4. Блог Open Source Vulnerability Database (OSVDB). Режим электронного доступа: https://blog.osvdb.org/ (дата обращения 23.07.2018).
5. Официальный веб-сайт Common Vulnerabilities and Exposures (CVE). Режим электронного доступа: https://cve.mitre.org/ (дата обращения 23.07.2018).
6. Официальный веб-сайт National Vulnerability Database (NVD). Режим электронного доступа: https://nvd.nist.gov (дата обращения 23.07.2018).
7. Официальный веб-сайт Common Attack Pattern Enumeration and Classification (CAPEC). Режим электронного доступа: https://capec.mitre.org/ (дата обращения 23.07.2018).
8. Официальный веб-сайт Common Weakness Enumeration (CWE). Режим электронного доступа: https://cwe.mitre.org/ (дата обращения 23.07.2018).
9. Официальный веб-сайт Offensive security's exploit database archive. Режим электронного доступа: https://www.exploit-db.com/ (дата обращения 23.07.2018).
10. Официальный веб-сайт Common Platform Enumeration (CPE). Режим электронного доступа: https://nvd.nist.gov/cpe.cfm (дата обращения 23.07.2018).
11. Официальный веб-сайт Common Configuration Enumeration (CCE). Режим электронного доступа: https://nvd.nist.gov /cce/index.cfm (дата обращения 23.07.2018).
12. Официальный веб-сайт Common Vulnerability Scoring System (CVSS). Режим электронного доступа Web: https://www.first.org/cvss (дата обращения 23.07.2018).

МЕТОДИКА ЗАЩИТЫ ДЕТЕЙ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ОТ НЕЖЕЛАТЕЛЬНОЙ ИНФОРМАЦИИ В СЕТИ ИНТЕРНЕТ

Э. Р. Хакимова[1]
(*Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, С.-Петербург, 197101, Кронверкский проспект, 49, elvirochka1995@mail.ru*)

А. А. Чечулин[2]
(*Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский институт информатики и автоматизации» Российской академии наук, С.-Петербург, 199178, 14-я линия В.О., 39, andreych@bk.ru*)

Аннотация

В настоящее время в области защиты Интернет-пользователей от нежелательной информации проводятся многочисленные исследования и разрабатываются новые методы и подходы для решения этой проблемы. Научный проект посвящен решению проблемы защиты детей с ограниченными возможностями от нежелательной информации в сети Интернет. Рассматриваются виды и источники нежелательной информации в сети Интернет, определена целевая аудитория для защиты от такой информации. проведен анализ методов машинного обучения и подходов к классификации веб-сайтов. Результатом является разработанная методика защиты детей с ограниченными возможностями от нежелательной информации в сети Интернет.

Ключевые слова: нежелательная информация, Интернет, Data Mining, защита от информации, категорирование веб-сайтов.

Введение

Предлагаемый доклад посвящен решению проблемы защиты Интернет-пользователей от нежелательной информации. В настоящее время происходит постоянный рост объема доступной информации, представленной в электронном виде. Не вся информация, содержащаяся в электронных источниках, является желательной к распространению.

Для защиты от нежелательного или запрещенного контента необходимо использовать эффективные механизмы, классифицирующие информацию и регулирующие доступ к ней. В первую очередь это важно для ограничения доступа к определенным видам информации по возрастным категориям [1, 2]. Особое внимание уделяется информации, запрещенной для изучения детьми, так как она может повлиять на их здоровье, а также физическое, психическое, духовное или нравственное развитие.

Таким образом, целью данной работы является защита детей с ограниченными возможностями от нежелательной или запрещенной информации в сети Интернет.

В рамках данного исследования ставятся следующие задачи:

1. Разработать классификацию электронных источников;
2. Разработать классификацию нежелательной информации для детей;
3. Провести сравнительный анализ методов классификации сайтов;
4. Провести сравнительный анализ методов машинного обучения;
5. Определить критерии оценки методов машинного обучения при классификации веб-сайтов;
6. Разработать методику защиты детей с ограниченными возможностями от нежелательной информации в сети Интернет.

Целевая аудитория системы защиты от нежелательной информации.

В настоящее время распространение нежелательной информации (НИ) является одной из важнейших проблем общества. К нежелательной информации можно отнести как запрещённую

[1] Магистр.
[2] Кандидат технических наук, ведущий научный сотрудник.

информацию, виды которой устанавливаются законодательством Российской Федерации (РФ), а также иную информацию, определяемую самим пользователем. Классификация целевой аудитории пользователей, подлежащих защите от НИ, представлена на рисунке 1.



Рис. 1. Целевая аудитория системы защиты от нежелательной информации

Особенно важно защитить от НИ детей, потому что получение такой информации может существенно повлиять на их дальнейшее развитие и психику. Также выбранной целевой аудиторией являются дети с ограниченными возможностями, находящиеся на инклюзивном обучении с использованием электронных устройств [3].

Классификация электронных источников в киберфизических системах

В настоящее время существует множество источников информации, как традиционных печатных (газеты, книги и т.д.), так и электронных. Классификация электронных источников информации представлена на рисунке 2 [4].



Рис. 2. Классификация электронных источников информации

Рассмотрим наиболее популярные из них. Электронная книга – это версия книги, хранящаяся в электронном виде. Данный термин применяется как для произведений, представленных в цифровой форме, так и в отношении устройств, используемых для их прочтения [5]. Также электронным источником информации может являться база данных. База данных представляет собой совокупность самостоятельных материалов, систематизированных таким образом, чтобы эти материалы могли быть найдены и обработаны с помощью электронной вычислительной машины [6]. Такой источник информации, как радио может применяться для оповещения населения о чрезвычайных ситуациях или транслирования новостей и музыки [7]. Также не менее популярным источником является телевидение. Основными электронными источниками ин-

формации в современном обществе являются сайты в сети Интернет. В таблице 1 представлено сравнение приведенных источников по способу отображения информации.

Таблица 1

Сравнительная характеристика электронных источников информации

Электронный источник информации	Способ отображения информации
Электронная книга	Текстовый, числовой, графический
База данных	Текстовый, числовой
Радио	Звуковой
Телевидение	Текстовый, видео
Веб-сайт	Текстовый, числовой, графический, звуковой, видео

Таким образом из таблицы 1 видно, что наиболее содержательными электронными источниками информации являются веб-сайты, на которых может быть размещена как необходимая пользователю информация, так и нежелательная, доступ к которой подлежит к ограничению.

Классификация нежелательной информации

В РФ имеется перечень информации, распространение которой запрещено. Он представлен в статье 15.1 Федерального закона N 149-ФЗ «Об информации, информационных технологиях и о защите информации» [8]. Данный вид запрещенной информации относится ко всем слоям населения, независимо от возрастных категорий и других признаков.

Для детей понятие запрещенной информации описано в законе N 436-ФЗ «О защите детей от информации, причиняющей вред их здоровью и развитию» [9]. Запрещенную или нежелательную к изучению детьми информацию определяет не только законодательство, но и контроль родителей, так как на них в первую очередь лежит ответственность за жизнь и здоровье детей. Основные категории нежелательной информации для детей представлены на рисунке 3.

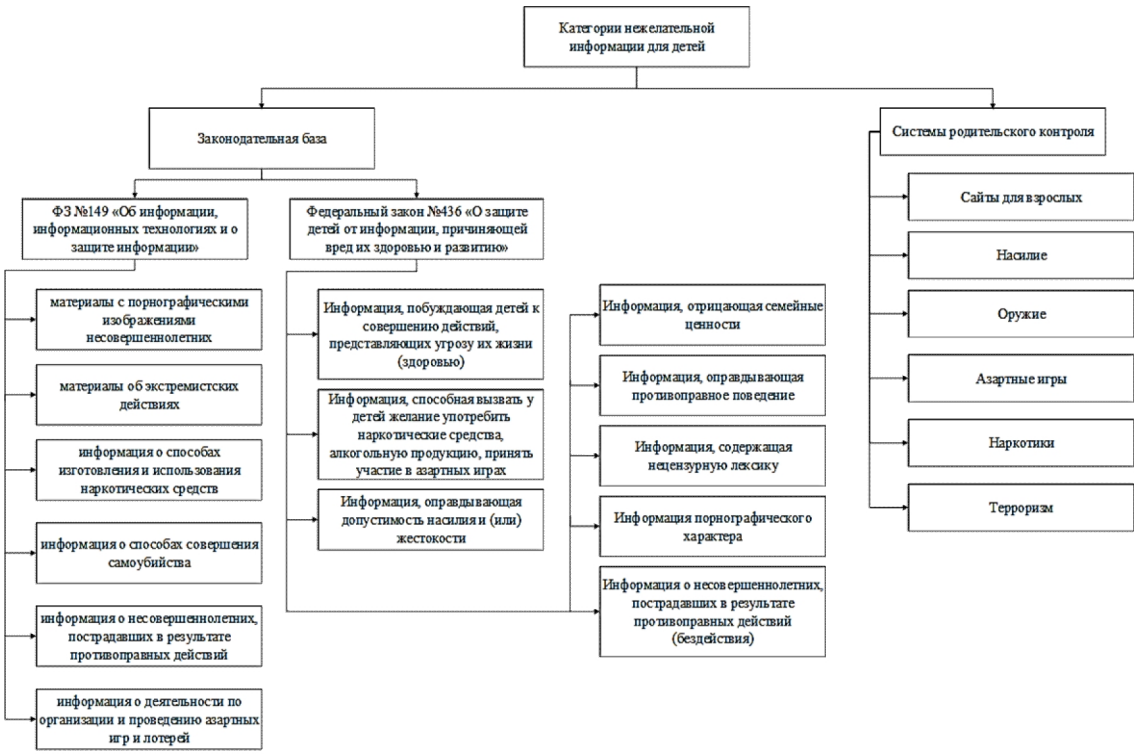


Рис. 3. Категории нежелательной информации для детей

С целью ограничения детей от нежелательного контента, который мог бы повлиять на их дальнейшее развитие и психическое здоровье, необходимо производить классификацию веб-сайтов, содержащих данную информацию для проведения дальнейшей фильтрации.

Источники данных о содержимом сайтов

Сайт в Интернете представляет собой массив связанных данных, имеющий уникальный адрес и воспринимаемый пользователем как единое целое. Страницы сайтов – это набор текстовых файлов, размеченных на языке HTML, позволяющий форматировать текст, различать в нём функциональные элементы, создавать гиперссылки и вставлять в отображаемую страницу изображения, звукозаписи и другие мультимедийные элементы [10].

Основными источниками данных о содержимом сайтов являются: URL-адрес, HTML-разметка, текстовое содержимое, графическая информация, звуковая информация, а также видео и медиафайлы.

URL (адрес сайта или отдельной страницы) – это специальная форма обозначения индивидуального адреса ресурса в интернете [11]. В состав HTML-разметки входит множество элементов: форматированное текстовое и графическое содержимое, программный код, ссылки на другие страницы, html-теги, которые также могут служить информацией о веб-сайте [12]. Основную смысловую нагрузку сайта составляет его текстовое содержимое, которое может дополняться медиаконтентом или графическими иллюстрациями, упрощающими восприятие информации пользователем.

Обзор методов классификации сайтов.

Существует множество различных подходов и методов к классификации сайтов. Критериями классификации могут выступать: возрастные ограничения, наличие запрещённой информации для пользователя, тематика сайта и другие. Основными подходами к классификации веб-сайтов являются: анализ URL-адреса, анализ HTML-тегов, анализ объектов на странице, не связанных с текстом (например, медиаконтент) или анализ текстового содержимого.

Классификация веб-страниц по URL-адресу может быть осуществлена при условии, что адрес отражает тематику веб-сайта. Можно анализировать URL-адрес целиком, а также можно анализировать его фрагменты с целью нахождения последовательностей символов, характерных для сайтов с нежелательной информацией. Главным недостатком данного метода является то, что не всегда URL-адрес отражает содержимое сайта [13].

Подход к классификации веб-сайтов, основанный на анализе HTML-тегов позволяет выделять из содержимой веб-страницы ключевые поля, такие как заголовки, названия разделов, подписи и др. Это позволяет не анализировать всю страницу целиком, а лишь её части, что существенно повышает скорость работы классификатора. Данный подход позволяет с высокой степенью точности отнести веб-страницу к классу, основанному на структурных признаках сайта, но не подходит для классификации, основанной на смысловых признаках.

Анализ изображений и другого контента, размещённого на веб-странице может сыграть очень важную роль в процессе выбора класса, которому принадлежит анализируемая страница, но далеко не всегда на веб-страницах размещённый контент соответствует её содержанию, поэтому предлагаемый метод может использоваться лишь как дополнение к тестовому анализу.

Для классификации веб-страницы по признаку возрастной принадлежности, а также наличию нежелательной или запрещённой для распространения информации наиболее эффективным и широко используемым является анализ текстового содержания веб-страниц. Анализ текста будет наиболее точным, так как текст отражает тематическую направленность веб-сайта и несет наибольшую смысловую нагрузку.

Обзор методов машинного обучения

Рассмотрим способы классификация текстового содержимого. Классификация может осуществляться полностью вручную, автоматически с помощью созданного набора правил, либо с применением методов машинного обучения.

Первым рассматриваемым способом является ручная классификация. Данная классификация осуществляется полностью человеком без средств автоматизации и как правило занимает дли-

тельное время. Таким образом, ручная классификация не всегда применима для обработки большого количества документов, если скорость обработки является главным показателем [13].

Второй подход заключается в написании правил, по которым можно отнести текст к категории. Этот подход лучше предыдущего, поскольку процесс классификации автоматизируется и, следовательно, количество обрабатываемых документов возрастает. Однако создание и поддержание правил в актуальном состоянии требует постоянных усилий специалиста.

Основывая на описанных ранних недостатках используем третий подход к классификации текстов основанный на машинном обучении (Data Mining). В этом подходе критерий принятия решения текстового классификатора, вычисляется автоматически из обучающих данных.

Data Mining – это собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. Одной из задач, решаемой методами Data Mining является классификация объектов (для заранее заданных классов), что поможет при классификации веб-страниц [14]. На рисунке 4 представлена классификация подходов к интеллектуальному анализу текстов.



Рис. 4. Классификация подходов к интеллектуальному анализу текстов

Большинство аналитических методов, используемые в технологии Data Mining – это известные математические алгоритмы и методы. Новым в их применении является возможность их использования при решении тех или иных конкретных проблем. Основные методы и алгоритмы Data Mining представлены на рисунке 5 [15].

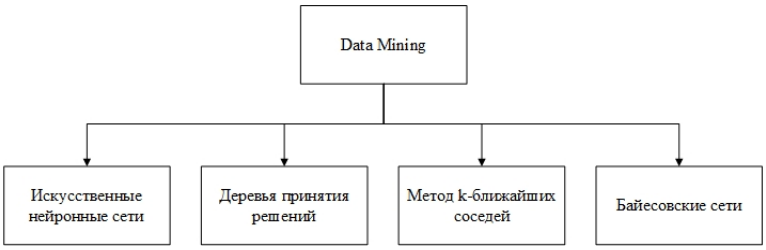


Рис. 5. Методы Data Mining

Рассмотрим представленные методы Data Mining для решения задачи классификации вебсайтов по текстовому содержимому. Первым рассматриваемым методом является нейронная сеть – это математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей [16]. С математической точки зрения нейронная сеть представляет собой граф с n вершинами-входами (текстовое содержимое сайтов) и k вершиной-выходами (категориями). Значение нейрона определяется весами ребер, входящими в него и значением нейронов на противоположных концах этих ребер. На рисунке 6 представлена структура нейронной сети.

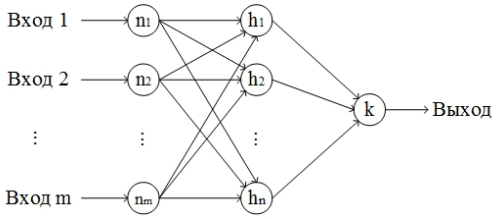


Рис. 6. Структура нейронной сети

На данном рисунке входные нейроны обозначаются как n_m и соответствуют словам, содержащимся на веб-сайте. Если слова веб-сайта относятся к одной категории из словаря, то они объединяются в группы и образуют нейроны из скрытого слоя h_n. Результатом работы нейронной сети является нейрон выходного слоя k, представляющий собой категорию, которой соответствует веб-сайт. Основными достоинствами данного метода являются: решение задач при неизвестных закономерностях, адаптирование к изменениям в окружающей среде, быстродействие, отказоустойчивость [17]. Несмотря на широкий спектр возможностей нейронных сетей, данный метод имеет ряд недостатков: отсутствие однозначных решений в большинстве случаев, продолжительные временные затраты на обучение, сложная реализация [18].

Вторым методом является метод деревьев решений. Деревья принятия решений – это иерархическая структура данных, которая способна не только решить задачу классификации, но и позволяет интерпретировать полученные результаты [13]. Деревья принятия решений строятся в виде графов, в которых узлами являются возможные события (категории), а ребра отображают их взаимосвязь. Конечным узлом является определенная категория конкретного веб-сайта. Каждому ребру в соответствие ставится вероятностная величина, определяющая дальнейшее ответвление дерева. К преимуществам метода можно отнести: быстрый процесс обучения, генерация правил в областях, где эксперту трудно формализовать свои знания и простота интерпретации. Среди недостатков: неприменимость к наборам данных, где число возможных исходов велико. Тогда деревья «переполнены данными», имеют много узлов и ветвей [19].

Третий рассматриваемый метод – метод k-ближайших соседей. Данный метод является достаточно простым и эффективным. Предсказание для новой точки делается путём поиска k-ближайших соседей в наборе данных и суммирования выходной переменной для этих k-экземпляров [20]. Данный алгоритм основывается на минимизации суммарного квадратичного отклонения точек классов от центров этих классов. В рассматриваемой задаче классификации веб-сайтов точкой класса является непосредственно веб-сайт, классом объекта является определяемая категория. Таким образом, сайт относится к той категории, ближе к которой расположен в векторном пространстве [21]. Алгоритм завершается, когда на какой-то итерации не происходит изменения внутреннего расстояния. Данный метод может потребовать много памяти для хранения всех данных, но зато быстро сделает предсказание. Также обучающие данные можно обновлять, чтобы предсказания оставались точными с течением времени [22].

Четвертым методом являются байесовские сети. Данный метод основывается на применении теоремы Байеса, известной из теории вероятности. Отличительной чертой является «наивное» предположение о независимости событий. Это сильное предположение, которое не соответствует реальным данным. Преимуществом алгоритма является малое количество данных для обучения, необходимых для оценки параметров при классификации, простота реализации и низкие вычислительные затраты. Анализ данных методов представлен в таблице 2.

Таблица 2

Анализ методов Data Mining

Название метода	Достоинства	Недостатки
Искусственные нейронные сети	– решение задач при неизвестных закономерностях – адаптирование к изменениям в окружающей среде – быстродействие – отказоучтойчивость	– отсутствие однозначных решений – не предсказуемость – продолжительные временные затраты – сложность схемотехнической реализации – сложность вычислительной реализации
Деревья принятия решений	– быстрый процесс обучения – генерация правил в трудных областях – простота интерпретации – высокая точность классификации	– неприменимость к наборам данных, где число возможных исходов велико – снижение показателя полноты данных.
k-ближайших соседей	– быстродействие – простота реализации – простота интерпретации	– неэффективный расход памяти – прямая зависимость количества данных и вычислительных операций
Байесовские сети	– малое количество данных – быстродействие – простота реализации – маленькие вычислительные затраты	– низкое качество классификации при реальных задачах

Таким образом, анализ текстового содержимого сайта будет производиться при помощи методов машинного обучения, так как для этого требуется наименьшее количество ресурсов. Метод деревьев решений является наиболее оптимальным методом для решения задачи классификации сайтов.

Определение критериев оценки методов машинного обучения при классификации веб-сайтов

Основой при оценке методов машинного обучения при классификации веб-сайтов является тестовая выборка. Чтобы принимать решение, хуже или лучше один классификатор работает, чем другой, необходима численная метрика оценки его качества. Критерии оценки методов машинного обучения представлены в таблице 3 [23].

Таблица 3

Критерии оценки методов машинного обучения

Название	Обозначение	Способ вычисления
Ошибка первого рода	FP	$FP = N_T - TP$
Ошибка второго рода	FN	$FN = N_F - TN$
Точность	P	$P = \frac{TP}{FP + TP}$
Полнота	R	$R = \frac{TP}{TP + FN}$
F-мера	F	$F = 2 * \frac{P * R}{P + R}$
Аккуратность	A	$A = \frac{\sum_{i=1}^{k}(TP_i + TN_i)}{\sum_{i=1}^{k} N}$
Время классификации	T	$T \rightarrow \min$

В данной таблице используются следующие обозначения: N_F – общее количество разрешенных сайтов; N_T – общее количество сайтов запрещенной категории; k – количество категорий; TP – количество однозначно запрещенных сайтов; TN – количество сайтов, однозначно принадлежащих разрешенной категории.

Ошибка первого рода (False Positive) – это ошибка, состоящая в опровержении верной гипотезы. Таким образом, это отнесение разрешенного веб-сайта к запрещенной категории. Ошибка второго рода (False Negative) – это ошибка, состоящая в принятии ложной гипотезы [24].

Количество сайтов, однозначно принадлежащих запрещенной категории (True Positive) определяется как количество запрещенных сайтов, которые система верно классифицировала. Количество однозначно разрешенных сайтов (True Negative) определяется как количество разрешенных сайтов, которые система верно классифицировала.

Точность (Precision) – это количество веб-сайтов, действительно принадлежащих запрещенной категории, относительно всех веб-сайтов, которые система отнесла к данной категории. Полнота (Recall) – это количество веб-сайтов, действительно принадлежащих запрещенной категории, относительно всех веб-сайтов этой категории. F-мера (F) – это метрика, объединяющая в себе информацию о точности и полноте, представляющая собой гармоническое среднее между этими двумя величинами. Из данного определения следует, что достижение максимальной точности и полноты одновременно невозможно.

Аккуратность (Accuracy) – это отношение количества веб-сайтов, по которым классификатор принял правильное решение, к общему количеству сайтов.

Время классификации (Time) – это время, затрачиваемое системой классификации при определении категории веб-сайта. Чем меньше времени затрачено на классификацию веб-сайта, тем лучше для пользователя. Время классификации определяется как время, затраченное системой на классификацию одинакового количества веб-сайтов различными методами машинного обучения. Данный параметр определяется экспериментально при одинаковых технических характеристиках устройства таких как: мощность процессора и оперативная память.

Методика защиты детей с ограниченными возможностями от нежелательной информации.

Для защиты детей от нежелательного контента в сети Интернет необходимо разработать методику, включающую в себя фильтрацию Интернет-трафика, а также дополнительные функции, позволяющие работать детям с ограниченными возможностями в сети Интернет. Разработанная методика представлена на рисунке 7.

Входной информацией для данной системы является Интернет-запрос, инициируемый ребенком, а выходной – отфильтрованный список разрешенных сайтов по заданному запросу.

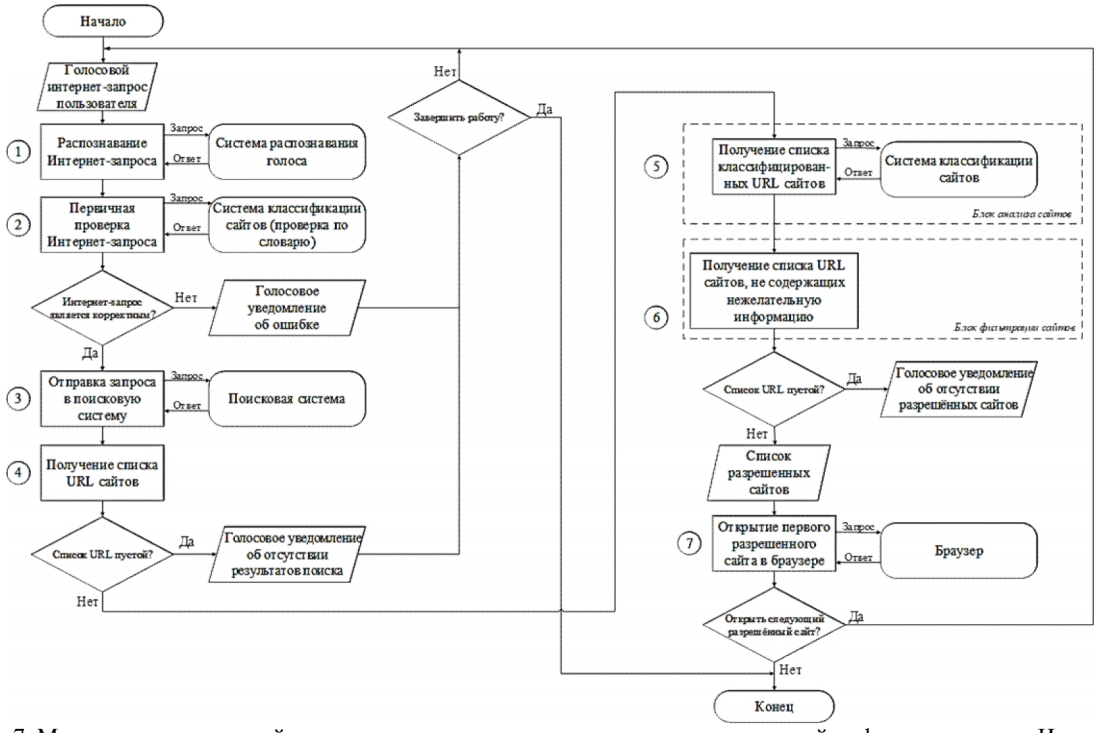


Рис. 7. Методика защиты детей с ограниченными возможностями от нежелательной информации в сети Интернет

(1) Так как система предназначена для детей с ограниченными возможностями, то ввод Интернет-запроса производится при помощи голосовых команд, а его обработка осуществляется системой распознавания голоса. (2) Перед тем, как перейти к классификации сайтов и ограничении доступа к ним необходимо проверить сам поисковой запрос пользователя. Если запрос уже содержит запрещенную информацию для детей, то поиск сайтов по данной тематике будет прекращен. (3) После первичной проверки запроса на наличие запрещенных слов данный запрос будет направлен в одну из поисковых систем сети Интернет. Выходными данными является файл, содержащий URL адреса сайтов по данному запросу, а также их описания и заголовки. (4) Из данного файла системе необходимы только URL адреса, по которым будет производиться дальнейшая классификация сайтов. Поэтому в результате производится разбор ответа поисковой системы и формируется список только URL адресов, относящихся к тематике, заданной пользователем. Если поисковая система ничего не обнаружила по заданному запросу, то пользователь будет уведомлен об этом и система перейдет в начальное состояние ожидания нового запроса. (5) Для получения списков классифицированный сайтов необходимо использовать машинно-обученный классификатор сайтов. (6) Для дальнейшей работы с имеющимися сайтами, необходимо отфильтровать те сайты, которые содержат нежелательную информацию. Для этого необходимо произвести фильтрацию полученного раннее списка и запретить доступ к тем сайтам, которые содержат НИ для детей. (7) Так как пользователями системы являются дети с ограниченными возможностями, то для удобства работы с данной системой будет производиться открытие первого разрешенного сайта по заданному запросу в браузере пользователя.

Таким образом, данная методика позволяет работать детям с ограниченными возможностями в сети Интернет, а также делает эту работу безопасной для здоровья и развития ребенка, ограничивая его от нежелательного контента.

Заключение

В настоящее время происходит постоянный рост объема доступной информации в электронном виде. Доступная информация не всегда является законной. В РФ имеется перечень информации, распространение которой строго запрещено. Но особое внимание уделяется информации, запрещенной для изучения детьми, так как она может повлиять на их дальнейшее развитие.

Наиболее содержательными электронными источниками информации в настоящее время являются веб-сайты. В связи с этим был разработан алгоритм классификации сайтов, содержащих запрещенную информацию для детей, с целью дальнейшего ограничения к ним доступа.

Для разработки алгоритма классификации были выявлены основные источники данных о содержимом сайтов, а также произведен анализ имеющихся методов классификации. Было выявлено, что анализ текстового содержимого сайтов будет наиболее точным, так как отражает его тематическую направленность, а производить данный анализ необходимо при помощи методов Data Mining.

Определены критерии оценки методов машинного обучения при классификации веб-сайтов. Основными критериями являются увеличение показателя точности, при снижении показателя полноты, а также время, затрачиваемое на классификацию веб-сайтов.

В связи с тем, что дети с ограниченными возможностями не всегда могут самостоятельно выбрать нужный сайт по заданной тематике, ввод поискового запроса осуществляется при помощи голосовых команд, а его обработка – системой распознавания голоса. Таким образом была разработана методика защиты детей с ограниченными возможностями от нежелательной информации в сети Интернет, позволяющая детям безопасно работать в сети, без риска для их здоровья и развития.

Работа выполнена при финансовой поддержке Гранта РНФ (проект № 18-11-00302) в СПИИРАН.

ЛИТЕРАТУРА

1. Kotenko I., Chechulin A., Komashinsky D. Categorisation of web pages for protection against inappropriate content in the internet // International Journal of Internet Protocol Technology (IJIPT). 2017. Vol. 10, no. 1. P. 61–71.
2. Котенко И.В., Чечулин А.А., Комашинский Д.В. Автоматизированное категорирование веб-сайтов для блокировки веб-страниц с неприемлемым содержимым // Проблемы информационной безопасности. Компьютерные системы. 2015. № 2. С. 62–68.
3. Назарова Н.М. Специальная педагогика: Учеб. пособие для студ. высш. учеб. заведений [Текст] / Л.И. Аксенова, Б.А. Архипов. – М.: Издательский центр «Академия». – 2005. – 400 с.
4. Абдуллаева Н.Д. Роль электронных библиотек и библиотечных инноваций в развитии информационного общества [Электронный ресурс] – Режим доступа: http://www.gpntb.ru/libcom15/tezis/010a.pdf (30.07.2018).
5. Stallman R. E-Books: Freedom Or Copyright [Электронный ресурс] – Режим доступа: https://www.gnu.org/philosophy/ebooks.en.html (30.07.2018).
6. Переводы, иные производные произведения. Составные произведения. [Электронный ресурс]: Статья 1260 Гражданского кодекса Российской Федерации от 18 декабря 2006 года N 230-ФЗ. – Режим доступа: http://www.consultant.ru/document/cons_doc_LAW_64629/26eaf5de7ca59025f4388fe2980d3dd03dd5e775/ (30.07.2018).
7. Радио обладает общественным воздействием и дает доступ к информации [Электронный ресурс] – Режим доступа: http://www.diamundialradio.org/2016/ru/theme/radio-obladaet-obsestvennym-vozdeistviem-i-daet-dostup-k-informacii/ (30.07.2018).
8. Об информации, информационных технологиях и о защите информации [Текст]: Федеральный закон от 27.07.2006 г. № 149-ФЗ (ред. от 25.11.2017).
9. О защите детей от информации, причиняющей вред их здоровью и развитию [Электронный ресурс]: федер. закон от 29.12.2010 №436-ФЗ (ред. от 29.06.2015). – Режим доступа: http://www.consultant.ru/document/cons_doc_LAW_108808/ (04.03.2018).
10. Ромашев В. CMS Drupal: Система управления содержимым сайта. – Питер. – 2010. – 255 с.
11. URL (URL-адрес). [Электронный ресурс] – Режим доступа: http://promo-sa.ru/seo-terms/URL-adress (30.07.2018).

12. Правила построения HTML-документов. [Электронный ресурс] – Режим доступа: http://www.webnav.ru/books/html4/html_rules/ (30.07.2018).
13. Новожилов, Д.А. Улучшение категорирования веб-сайтов для блокировки неприемлемого содержимого на основе анализа статистики html-тэгов [Текст] / Д.А. Новожилов, А.А. Чечулин, И.В. Котенко. – Спб.: Информационно управляющие системы. – 2016. – №6.
14. Дюк В. Data Mining: учебный курс. [Текст] / Дюк В., Самойленко А. – СПб.: Изд. Питер, 2001. – 368 с.
15. Лепшова Е. Обзор методов Data Mining. [Электронный ресурс] – Режим доступа: http://intellect-tver.ru/?p=165 (30.07.2018).
16. Осовский С. Нейронные сети для обработки информации. [Текст] / С. Осовский. – М.: Финансы и статистика, 2002. – 344 с.
17. Преимущества нейронных сетей. [Электронный ресурс] – Режим доступа: http://www.aiportal.ru/articles/neural-networks/advantages.html (30.07.2018).
18. Основные недостатки использования искусственных нейронных сетей и пути их решения. [Электронный ресурс] – Режим доступа: http://neuronus.com/theory/242-osnovnye-nedostatki-ispolzovaniya-iskusstvennykh-nejronnykh-setej-i-puti-ikh-resheniya.html (30.07.2018).
19. Левитин А. В. Ограничения мощи алгоритмов: Деревья принятия решения. [Текст] / А.В. Левитин. – М.: Вильямс. – 2006. – 576 с.
20. Le J. A Tour of The Top 10 Algorithms for Machine Learning Newbies. [Электронный ресурс] – Режим доступа: https://towardsdatascience.com/a-tour-of-the-top-10-algorithms-for-machine-learning-newbies-dde4edffae11 (30.07.2018)
21. Лавров Д.Н. Применение кластерного анализа методом k-средних для классификации текстов научной направленности. [Текст] / Д.Н. Лавров, Ю.А. Осипова. – Омск: Омский государственный университет. – 2017. – 116 с.
22. Wang H. Extended k-Nearest Neighbours Based on Evidence Theory. [Текст] / H. Wang, D. Bell. – The Computer Journal. – 2004. – 672 с.
23. Баженов Д. Оценка классификатора (точность, полнота, F-мера) [Электронный ресурс] – Режим доступа: http://bazhenov.me/blog/2012/07/21/classification-performance-evaluation.html (230.07.2018)
24. Яковлева А. Ошибки первого и второго рода. Понятие о статистических критериях. [Электронный ресурс] – Режим доступа: https://www.e-reading.club/chapter.php/1002275/20/Yakovleva_Angelina.html (230.07.2018).

E.R. Khakimova (Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics, Saint-Petersburg), **A.A. Chechulin** (Saint-Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, Saint-Petersburg)

Protection technique against inappropriate information in the internet network for children with disabilities

Currently, in the field of protecting Internet users from inappropriate information, numerous studies are being conducted and new methods and approaches are being developed to solve this problem. The investigation is devoted to solving the problem of children with disabilities protection against inappropriate information in the Internet. The types and sources of inappropriate information in the Internet are considered, and the target audience is identified. An analysis of methods of machine learning and approaches to the classification of websites is performed. The result is the developed method of children with disabilities protection from inappropriate information in the Internet.

МЕТОДИКА ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ СЕТЕВОГО ТРАФИКА ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛИЙ В КИБЕРФИЗИЧЕСКИХ СИСТЕМАХ

Д. А. Гайфулина
(*Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики (Университет ИТМО),*
С.-Петербург, Кронверкский пр., 49, wing7803@yandex.ru)

А. В. Федорченко
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), С.-Петербург, 14-я линия В.О., 39*
fedorchenko@comsec.spb.ru)

Аннотация

Работа посвящена определению нерегламентированных сетевых протоколов взаимодействия в сетевом трафике киберфизических систем для повышения эффективности применения методов обнаружения аномалий. Предлагаемый подход основан на типовом и структурном определении сетевых протоколов за счет обработки трафика методами структурного анализа. Приводится модель сетевого трафика и методика его предварительной обработки. Описываются основные этапы реализации методики, а также формулируются требования к исходным данным. Экспериментальная оценка точности предлагаемого подхода подтверждает возможность его применения для задачи выявления типовых структур пакетов в трафике с нерегламентированными сетевыми протоколами взаимодействия.

Ключевые слова: сетевой трафик, протоколы передачи информации, структурный анализ, информационная безопасность.

Введение

В современном мире киберфизические системы (КФС) становятся все более востребованными, так как позволяют осуществлять интеграцию информационных технологий и устройств взаимодействия с физическими процессами и объектами. [1].Одними из наиболее часто используемых сетевых интерфейсов в КФС являются интерфейсы взаимодействия между микроконтроллерами и внешними устройствами [2, 3]. Таким образом, передача информации осуществляется преимущественно на низких уровнях взаимодействия, осуществляющих передачу данных по физическим каналам связи. Передача информации может осуществляться как за счет оригинальных сетевых протоколов прикладного уровня, которые не всегда регламентированы, так и с использованием известных протоколов. По этой причине часть сетевого трафика может быть синтаксически и семантически не определенной.

Для обеспечения безопасности КФС используются различные методы оценки их защищенности[4-6]. Одним из таких методов является анализ сетевой активности с целью обнаружения аномалий – ситуаций, в которых поведение инфраструктуры системы или ее отдельных элементов отличается от эталонного поведения. Исходными данными для анализа в данном случае служит ряд признаков, характеризующих функционирование сетевой инфраструктуры, например, адреса абонентов, флаги сообщений, запросы соединений и др. [7]. В свою очередь, в условиях нерегламентированных протоколов сетевого взаимодействия (НПСВ) применение методик обнаружения сетевых аномалий сильно ограничено и требует предварительного анализа сетевого трафика. Идентификация параметров нерегламентированных сетевых протоколов для построения модели поведения системы и оценка ее защищенности является достаточно трудновыполнимой задачей, которая требует определенных сведений о сетевых протоколах. Таким образом, *целью* проводимого исследования является определение структур нерегламентированных протоколов сетевого взаимодействия (НПСВ)в КФС для повышения эффективности применения методов обнаружения сетевых аномалий.

Задачами исследования являются: (1) разработка методики выявления типовых структур пакетов сетевого трафика; (2) определение требований к задаче определения сетевых протоколов; (3) оценка точности методики предварительной обработки сетевого трафика в условиях НПСВ. Научная новизна проводимого исследования заключается в типовом и структурном определении

протоколов передачи информации в сетях киберфизических систем за счет обработки передаваемого сетевого трафика методами структурного анализа. Предлагаемый доклад посвящен разработке методики предварительной обработки трафика для определения нерегламентированных протоколов сетевого взаимодействия в КФС.

Разработка методики предварительной обработки сетевого трафика в условиях НПСВ

Выявление типовых структур пакетов сетевого трафика. В разрабатываемом подходе предполагается, что протоколы сетевого взаимодействия описываются различными типовыми структурами, представляющими собой упорядоченные наборы типовых последовательностей байт (бит) сетевого трафика.

Исходными данными предлагаемой методики является множество записей о полученных пакетах сетевого трафика, представляющее собой последовательность байт (бит) *S*.Выходными данными методики *Y* являются словарь типовых последовательностей («слов») сетевого трафика *W*, и словарь типовых структур для исходных данных сетевого трафика *PT*:

$$Y(S) = \{W, PT\},$$

где$S= \{s_1,s_2,..., s_n\}$ – последовательность данных (байт или бит информации) сетевого трафика, n – количество данных (в байтах или битах);$W =\{ w_1, w_2, \ldots, w_m\}$.– словарь типовых последовательностей байт (бит); $PT = \{pt_1, pt_2, .., pt_r\}$– словарь типовых структур сетевого трафика, а r – их общее число.

Для получения данного вектора выходных значений по заданным входным параметрам необходимо выполнение следующей последовательности операций.

1. Нахождение типовых последовательностей («слов»).

Операция выполняет разбиение исходного ряда данных на отрезки так, что запись сетевого трафика принимает вид:

$$S = \{d_0, d_1, ..., d_e\},\ d_i = \{s_{ik}, s_{ik+1}, ..., s_{ik+k}\}\ i \in (0,1,...,\ e-1),\ 2 \le k \le n-1$$

где d_i– последовательность байт (бит) сетевого трафика;e – число отрезков, k– число байт (бит) каждого отрезка.

Далее находятся все отрезкиd_i, повторяющиеся более одного раза среди подобных отрезков для каждого из возможных k. Итерации продолжаются до некоторого значения длины отрезка k, при котором в исходной последовательности данных сетевого трафика не будет найдено ни одного идентичного отрезка той же длины.Выходными данными является множество «слов»W: $W = \{w_1, w_2, ..., w_m\}$, где w – «слово»; m – количество выявленных «слов».

Каждое «слово» wописывается его численным индексом iw,значением vw, длиной k, и показателем встречаемости cw:

$$w = \{iw, vw, k,\ cw\},$$

$$IW = \{iw_1, iw_2, ..., iw_m\}, VW = \{vw^{iw_1}, vw^{iw_2}, ..., vw^{iw_m}\},\ K = \{k_1, k_2, ..., k_l\},\ CW = \{cw_1, cw_2, ..., cw_h\},$$

где IW – множество индексов «слов»; VW – множество значений «слов»; K – множество длин «слов», l – количество возможных длин «слов»; CW – множество показателей встречаемости «слов»; h – количество показателей встречаемости «слов».

2. Упорядочивание индексов типовых последовательностей

Далее определяются порядковые номера байт (бит) начала и конца вхождения «слова» в сетевом трафике. Участки сетевого трафика, в которых не присутствуют типовые последователь-

ности, соответствуют нулевому индексу, обозначающему типовую неопределенность подобных участков данных.

Выходными данными является последовательность индексов «слов» *I:*

$$I = \{iw'_1, iw'_2, ..., iw'_t\},$$

$$iw'_t = \begin{cases} iw_x,\ s_i \in w_x, w_y,\ if\ k^x cw^x > k^y cw^y, \\ iw_y,\ s_i \in w_x, w_y,\ if\ k^x cw^x < k^y cw^y, \\ 0,\ s_i \notin w_x, w_y \end{cases} \quad w_x, w_y \in W,\ j = (1,2,...,t)$$

гдеs_i– бит (байт) сетевого трафика;iw_xиiw_y– индексы «слов»w^xи w^yс длинами k^x, k^yи показателями встречаемости cw^x, cw^y соответственно; t– длина последовательности индексов.

3. Определение типовых структур сетевого трафика

Список индексов *I*, аналогично исходным данным, разбивается на отрезки, представляющие собой упорядоченные наборы индексов, так что запись принимает вид:

$$I = \{di_0, di_1, ..., di_g\},\ di_i = \{iw'_{iz}, iw'_{iz+1}, ..., iw'_{iz+z}\}\ i \in (0,1,...,\ g-1),\ z \geq 2,\ iw'_0 \neq 0,$$

где di_i– последовательность индексов «слов» (структура сетевого трафика); g – их общее число, z– число ненулевых индексов «слов» в каждомотрезке.

Затем производится вычисление коэффициента корреляции Пирсона C_{xy}между всеми парами отрезков di_x и di_y. Выбираются все структуры di_xдля которых выполняется условие $C_{xy} = 1$.Итерации продолжаются до некоторого значения длины последовательностей индексов, при котором в исходных данных не будет найдено ни одной пары последовательностей с коэффициентом корреляции равным 1, то есть $z = 0$.

Выходными данными является множество типовых структур*PT*:

$$PT = \{pt_1, pt_2, ..., pt_r\},$$

где pt – типовая структура; r – количество выявленных структур.Каждое структура ptописывается численным индексом ipt, значением vp, длиной z, и показателем встречаемости cp:

$$pt = \{ipt,\ vp,\ z,\ cp\},$$
$$IPT = \{ipt_1, ipt_2, ..., ipt_r\}, VW = \{vp^{ipt_1}, vp^{ipt_2}, ..., vp^{ipt_r}\}, vp^{ipt} \subset IW$$
$$Z = \{z_1, z_2, ..., z_q\},\ CP = \{cp_1, cp_2, ..., cp_h\},$$

где *IPT*– множество индексов структур; *VP* – множество наборов «слов»; *Z* – множество длин структур, q – количество возможных длин структур; *CP* – множество показателей встречаемости структур; b – количество показателей встречаемости структур.

С учетом разработанной модели сетевого трафика и методики его предварительной обработки в условиях НПСВ, были сформулированы требования к исходным данным по: (1) содержательности; (2) упорядоченности и (3) схожести структур заголовков пакетов сетевых протоколов.

Оценка точности методики предварительной обработки сетевого трафика. Предлагаемый подход анализа сетевого трафика для выявления НПСВ был практически реализован и апробирован на различных наборах исходных данных. Для оценки точности разрабатываемой методики на основе полученных значений проведены расчеты коэффициентов, представленных в таблице 1. Коэффициент покрытия представляет собой отношение совокупной длины выявленных «слов» и выявленных структур к общей длине трафика отдельно.

Таблица 1

Характеристика структуры записей сетевого трафика по результатам методики

Имя файла	Длина (байт)	Количество полей протоколов			Количество «слов»		Количество типовых структур	Коэффициент покрытия		*FWR = FR/FW*	*FSR = FS/FR*
		Общее	Неуник.	*FR = HR/H*	Общее	В покрытии		«словами»	типовыми структурами		
	N	*H*	*HR*		*MW*	*MWF*	*MS*	*FW*	*FS*		
dump_arp	930	13	13	1	241	17	14	0,9214	0,815	0,921	0,815
dump_tcp	1 129	23	19	0,952	226	39	21	0,901	0,883	0,946	0,927
dump_dhcp	1500	21	20	0,865	542	21	10	0,813	0,664	0,939	0,767
dump_http	2091	32	25	0,812	953	144	44	0,781	0,621	0,961	0,795
dump_dns	2 388	25	20	0,826	560	89	20	0,7713	0,6867	0,933	0,831
dump_udp	1557	52	45	0,8	795	66	31	0,762	0,587	0,952	0,733
Среднее значение по методике										**0,942**	**0,812**

Пример визуального представления байтовой структуры записи сетевого трафика в виде тепловой карты в оригинальном порядке полей протоколов представлен на рисунке 1-а. Аналогичным способом отображена разметка сетевого трафика по «словам» – рисунок 1-б, и по типовым структурам – рисунок 1-в. Нумерация байта производится по его принадлежности к определенному полю протокола, «слову» или типовой структуре в соответствии со словарем из каждой категорией.

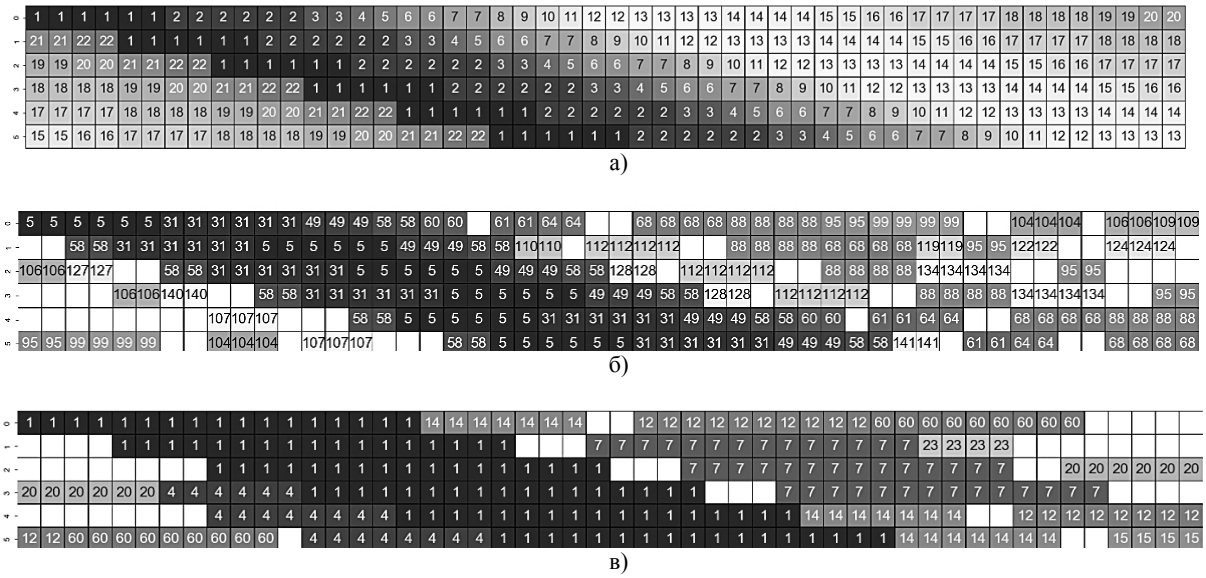


Рис. 1. Разметка сетевого трафика а) в оригинальном порядке полей протоколов, б) по «словам», в) по типовым структурам

На данных изображениях можно проследить схожесть определения «слов» сетевого трафика с положениями реальных полей протоколов. Также данные «слова» образуют повторяющиеся последоватльности, принимаемые за типовую структуру сетевого пакета. Таким образом, экспериментально показано, что разработанная методика позволяет определять типовые структуры в данных сетевого трафика с высокими показателями покрытия относительно оригинальных структур протоколов.

Заключение

В результате проведенных исследований разработана модель сетевого трафика КФС и методика его предварительной обработки, направленная на определение структур нерегламентиро-

ванных протоколов сетевого взаимодействия. Определены особенности структуры сетевого трафика КФС, по которым сделан вывод о ее возможной неопределенности. Определены требования к исходным данным, ограничивающие применение предлагаемого подхода выявления типовых структур сетевого трафика. По результатам экспериментальной оценки точности разработанной методики выявленные типовые последовательности по средним показателям на 94,2% соответствуют содержанию полей сетевых протоколов трафика, а выявленные структуры на 81,2% соответствуют спецификациям данных протоколов.

Практическая значимость результатов работы заключается в возможности выполнения предварительной обработки данных сетевого трафика для решения задачи обнаружения аномалий и исследования инцидентов информационной безопасности в условиях неопределенной сетевой инфраструктуры.

Работа выполнена при поддержке гранта президента РФ (МК-314.2017.9).

ЛИТЕРАТУРА

1. Henzinger T., Sifakis J. The Embedded Systems Design Challenge // Lecture Notes in Computer Science. 2006. Vol.4085. P. 1-15.
2. Денисенко, В. В. Компьютерное управление технологическим процессом, экспериментом, оборудованием // М.: Горячая линия–Телеком, 2014. – 608 с.
3. LeeE. A. ComputingFoundationsandPracticeforCyber-PhysicalSystems: APreliminaryReport // TechnicalReportUCB/EECS-2007-72, UniversityofCalifornia, Berkeleyб, 2007 – С 23-34.
4. Браницкий А.А., Котенко И.В. Анализ и классификация методов обнаружения сетевых атак системах // Труды СПИИРАН. 2016. Вып. 2(45). С.207-244.
5. Дойникова Е. В. Показатели и методики оценки защищенности компьютерных сетей на основе графов атак и графов зависимостей сервисов // Труды СПИИРАН. Вып.3 (26). СПб.: Наука, 2013. С.54-68.
6. Федорченко А.В., Левшун Д.С., Чечулин А.А., Котенко И.В. Анализ методов корреляции событий безопасности в SIEM-системах.Часть 2 // Труды СПИИРАН. 2016. Вып. 6(49). С.208-225.
7. Шелухин О.И., Симонян А.Г., Ванюшина А.В. Влияние структуры обучающей выборки на эффективность классификации приложений трафика методами машинного обучения // T-Comm: Телекоммуникации и транспорт. 2017. Том 11. №2. С. 25-31.

D.A. Gaifulina (University ITMO, St. Petersburg), **A.V. Fedorchenko** (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science, St. Petersburg)

Method of preliminary processing of network traffic for detecting anomalies in cyber-physical systems

The work is devoted to the definition of unregulated network protocols of interaction in network traffic of cyber-physical systems to improve the efficiency of the application of methods for detecting anomalies. The proposed approach based on the typical and structural definition of network protocols by processing network traffic using structural analysis methods.The model of network traffic and the method of its preliminary processing are given. The main stages of the implementation of the methodology are described, and the requirements to the initial data are formulated. The experimental estimation of the accuracy of the proposed approach confirms the possibility of its application for the task of identifying typical packet structures in network traffic with unregulated network protocols of interaction.

МЕХАНИЗМЫ ОБЕСПЕЧЕНИЯ ЦЕЛОСТНОСТИ ДАННЫХ В ОБЛАЧНЫХ СИСТЕМАХ, ОСНОВАННЫЕ НА ПРОТОКОЛАХ ДОКАЗАТЕЛЬСТВА ОБЛАДАНИЯ ДАННЫМИ

Е. С. Меркушев[1]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук, С.-Петербург, 14 линия В.О., д. 39, eugenemerkushev@gmail.com*)

Аннотация

В докладе проводится систематизация механизмов обеспечения целостности данных в облачных системах: выделяются наиболее распространенные угрозы, обсуждаются стратегии защиты, формулируются перспективные направления исследований в этой области.

Ключевые слова: целостность данных, облачные системы, протокол обладания данными

Введение

Одними из важнейших механизмов защиты информации являются механизмы обеспечения целостности [1]. Понятие целостность применительно к облачным системам подразумевает, что провайдер должен добросовестно хранить доверенные ему данные, а вычислительные задачи пользователей не должны подвергаться каким-либо воздействиям со стороны вредоносных утилит, других пользователей или самого облачного провайдера, причем любые нарушения при хранении данных в облаке (потеря, изменение, компрометация), а также любые некорректные вычисления должны быть обнаружены. Целостность остается одним из наиболее актуальных аспектов в сфере облачных технологий, поскольку в этом случае пользовательские данные и их вычислительные задачи хранятся удаленно на подконтрольных потенциально ненадежным провайдерам серверах, а традиционные механизмы обеспечения целостности зачастую оказываются неприменимы.

Основные угрозы целостности в облачных окружениях принято разделять в соответствии с определением, данным выше. Во-первых, это потеря, изменение или компрометация данных. На серверах провайдера хранится значительное количество информации пользователей. Часть данных может быть повреждена злонамеренно или случайно. Известны случаи [2], когда ошибки администраторов при резервном копировании, восстановлении или миграции приводили к потере данных. Кроме того, атаки различного рода со стороны третьих лиц так же могут привести к утечке, порче или потере данных. Во-вторых, некорректные облачные вычисления. Поскольку вычисления производятся на удаленных ресурсах провайдеров, механизмы вычислений недостаточно прозрачны для клиентов облачной инфраструктуры [3-5]. Сервера могут работать некорректно и возвращать ошибочный результат по различным причинам, среди которых неправильное конфигурирование провайдером программной или аппаратной части, наличие устаревшего и (или) уязвимого кода, присутствие зараженных вычислительных узлов [6-10].

Ввиду увеличения объема хранимой в облачных системах информации, традиционные методы, требующие вычисления хэша целого файла на стороне провайдера, становятся недоступными. Неоправданным оказывается и механизм, при котором необходимо скачивать файл с сервера, а затем самостоятельно выполнять проверку целостности, поскольку это влечет за собой большие вычислительные расходы и значительное потребление пропускной способности.

Целью работы является систематизация механизмов обеспечения целостности удаленно хранящихся данных, использующих протоколы доказательства обладания данными, обзор наиболее распространенных реализаций, формулировка перспективных направлений в этой области.

Протоколы доказательства обладания данными

Существует несколько подходов к обеспечению целостности удаленно хранящихся данных. Наиболее распространенными являются механизмы, основанные на протоколе доказательства

[1] Аспирант.

возможности возврата [11], протоколе доказательства обладания данными [2,12], модели слоя высокой доступности и целостности для облачного хранения [13], привлечении стороннего аудитора [14].

В контексте настоящего доклада рассматриваются различные реализации протокола доказательства обладания данными (Proof of Data Possession). Рассматриваемый протокол предполагает хранение метаданных на стороне клиента. Процесс проверки целостности разделен на четыре фазы: подготовительная – вычисление метаданных, которые должны храниться локально; пользовательский запрос проверки состояния файла/данных, хранящихся удаленно; ответ сервера; проверка ответа на стороне клиента.

Простейшая реализация протокола обладания данными заключается в предварительном вычислении значений одной или нескольких хэш-функций на файле, который после этого перемещается на сервер. Полученные значения хэш-функций являются метаданными и хранятся локально. Для проверки целостности файла клиент отправляет запрос, при получении которого сервер вычисляет хэш хранящегося удаленно файла и отправляет его пользователю. В случае, если имеющиеся у клиента данные совпадают с полученными, утверждается, что целостность не нарушена. Описанная реализация является вычислительно затратной даже для простых операций хэширования, крайне редко применяется в облачных системах хранения данных.

Оригинальный протокол доказательства обладания данными лишен недостатка простейшей реализации. Подготовительная фаза заключается в разбиении файла на конечное число покрывающих его, попарно непересекающихся подмножеств – блоков. На стороне клиента генерируется пара ключей: публичный и секретный. Для блоков с использованием ключей вычисляются гомоморфные, неподдельные теги, размер которых должен быть существенно меньше соответствующих блоков. Исходный файл в виде множества блоков вместе с созданными тегами и публичным ключом отправляется на сервер [2]. Рассматриваемый подход не является вычислительно затратным, поскольку калькуляция значений функций полностью на всем файле заменена аналогичной операцией на нескольких блоках. Для проверки целостности клиент отправляет на сервер непустое множество номеров блоков. Удаленное хранилище, используя имеющиеся данные (публичный ключ, тег и блок файла) генерирует доказательство. Процесс выполняется для всех блоков, запрашиваемых пользователем, после чего набор доказательств, тегов и блоков исходного файла, соответствующих запросу клиента, отправляются пользователю. Используя приватный ключ и полученные данные, владелец проверяет подлинность каждого запрошенного им блока. Описанная модель позволяет достичь высокой вероятности определения некорректной работы сервера при незначительных вычислительных расходах и издержках на хранение информации, однако ее применимость ограничена хранением неизменяющихся после загрузки на сервер данных.

Динамическая модель протокола доказательства обладания данными поддерживает модификацию данных [12]. Данная реализация имеет три новые операции: «подготовить изменение», «применить изменение», «проверить изменение». Первая запускается клиентом для подготовки запроса на обновление данных (изменить блок, удалить блок и т.д.). Вторая запускается сервером для фактического обновления файла, после чего возвращает доказательство изменения клиенту, который, в свою очередь, проверяет корректность проведенных на сервере изменений данных.

Архитектура оригинального протокола формально позволяет реализовать операции добавления, замены и удаления блоков, однако без введения дополнительных структур данных невозможно определить версионность блоков, хранящихся на сервере, поскольку как актуальные, так и устаревшие данные в классическом подходе будут иметь валидные теги. Для учета версионности в динамической реализации используется бинарное дерево, именуемое «ранговый список подлинных переходов». Количество листьев равно мощности множества блоков данных. Листовые узлы дерева всегда указывают на актуальные версии блоков. Каждый узел хранит два параметра: натуральное число, равное количеству блоков, содержащихся в поддереве, корневым элементом которого является рассматриваемый узел, и значение определенной функции (для простоты – хэш-функции), зависящее от аналогичного параметра наследников. Листья, таким образом, содержат единицу в качестве первого параметра и значение хэш-функции тега того блока, с которым они непосредственно связаны, в качестве второго. Любые операции мо-

дификации файла приводят к пересчету параметров тех узлов, из которых можно прийти к изменяемым блокам. Особое внимание уделяется корневому элементу дерева, поскольку его параметры изменяются при любой модификации данных. Значение хэш-функции корневого узла в динамической модели позволяет определять целостность файла после его модификации. Результаты экспериментов показывают, что, несмотря на рост вычислительных затрат, модель оказывается эффективной. Например, для проверки целостности файла размером 1Гб, динамической модели необходимо сгенерировать всего 415Кб данных и потратить 30мс на вычислительные расходы.

Недостатком данного подхода является отсутствие возможности публичной проверки целостности, необходимость в которой возникает, если данные, хранящиеся удаленно, становятся общими для группы пользователей.

Перспективным направлением в области обеспечения целостности хранимых удаленно данных является разработка эффективного механизма (или модификация имеющихся) с целью удовлетворить ряду критериев: возможность как публичной, так и приватной проверки целостности; низкие вычислительные затраты; слабые требования на ширину полосы пропускания; возможность неограниченного количества проверок и их независимость.

Заключение

Касательно целостности данных в облачных системах выделяют две основные проблемы: значительные объемы данных исключают использование обычных алгоритмов хэширования; проверка целостности может быть применена только после реализации дополнительных требований, которые увеличивают сложность. В работе рассмотрены механизмы обеспечения целостности данных в облачных системах, использующие протоколы доказательства обладания данными, выделены преимущества и недостатки наиболее распространенных реализаций, на основании изложенного материала сформулированы перспективные направления в этой области.

Авторами в настоящее время ведутся исследования по защите информации при реализации облачных технологий больших данных для мониторинга безопасности [15, 16].

Работа выполняется при финансовой поддержке РФФИ (проекты 16-29-09482 и 18-07-01488) и бюджетной темы АААА-А16-116033110102-5.

ЛИТЕРАТУРА

1. Котенко И.В., Юсупов Р.М. Перспективные направления исследований в области компьютерной безопасности //Защита информации.Инсайд. 2006. №2. С.46-57.
2. Ateniese G., Burns R., Curtmola R., Herring J., Kissner L., Peterson Z., Song D. Provable data possession at untrusted stores // ACM conference on Computer and communications security, ACM, 2007. P.598–609.
3. Xiao Z., Xiao Y. Accountable MapReduce in Cloud Computing // IEEE Conference on Computer Communications Workshops. IEEE, 2011. P.1082–1087.
4. Haeberlen A., Kuznetsov P., Druschel P. PeerReview: Practical accountability for distributed systems // ACM Symposium on Operating Systems Principles. 2007. P.175-188.
5. Santos N., Gummadi K. P., Rodrigues R. Towards trusted cloud computing // Conference on hot topics in cloud computing. USENIX Association Berkeley, 2009.
6. Десницкий В.А., Котенко И.В. Защита программного обеспечения на основе механизма “удаленного доверия”//Изв.вузов.Приборостроение,Т.51,№11,2008. С.26-30.
7. Десницкий В.А., Котенко И.В. Методы защиты программного обеспечения на основе принципа удаленного доверия //Защита информации.Инсайд,№ 6, 2009. С.57-61.
8. Котенко И.В., Десницкий В.А. Аспектно-ориентированная реализация модели защиты программ на основе “удаленного доверия” // Информационные технологии и вычислительные системы, № 4, 2009, С.67-76.
9. Wang C., Ren K., Wang J. Secure and Practical Outsourcing of Linear Programming in Cloud Computing // IEEE Trans. Cloud Computing: IEEE, 2011. P.820–828.
10. Monrose F., Wyckoff P., Rubin A. D. Distributed execution with remote audit // Network and Distributed System Security Symposium. ISOC, 1999.
11. Juels A., Kaliski B. S. PORs: Proofs of retrievability for large files // ACM conference on Computer and communications security. ACM, 2007. P. 584–597.

12. Erway C., Küpçü A., Papamanthou C., Tamassia R. Dynamic provable data possession // ACM conference on Computer and communications security. 2009. P.213–222.
13. Bowers K. D., Juels A., Oprea A. HAIL: A high-availability and integrity layer for cloud storage //ACM conference on Computer and communications security,2009. P.187-198.
14. Wang C., Wang Q., Ren K., Lou W. Privacy-preserving public auditing for data storage security in cloud computing // IEEE INFOCOM 2010, IEEE, 2010. P. 1–9.
15. Novikova E., Kotenko I. Analytical Visualization Techniques for Security Information and Event Management // Proceedings of the 2013 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, PDP 2013. P.519-525.
16. Котенко И. В., Полубелова О.В., Саенко И.Б., Чечулин А.А. Применение онтологий и логического вывода для управления информацией и событиями безопасности // Системы высокой доступности. 2012. Т. 8. № 2. С. 100-108.

E.S. Merkushev (Saint-Petersburg university to informatics and automation of Russian Academy of Sciences, Saint-Petersburg)

Cloud data integrity mechanisms based on proof of data possession protocol

The article is concerned with the existing data integrity issues in cloud environments. Vulnerabilities that can be exploited by attackers, the threat models, as well as existing defense strategies in a cloud scenario are discussed. It gives a comprehensive review of data integrity mechanisms based on proof of data possession protocols.

ВИЗУАЛИЗАЦИЯ ДАННЫХ В ИНТЕРЕСАХ ВЫЯВЛЕНИЯ И ПРОТИВОДЕЙСТВИЯ НЕЖЕЛАТЕЛЬНОЙ И СОМНИТЕЛЬНОЙ ИНФОРМАЦИИ

М. В. Коломеец[1], **Д. С. Левшун**[2]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), 199178, Россия, С.-Петербург, 14-я линия В.О., д. 39, (812) 328-71-81, kolomeec@comsec.spb.ru, levshun@comsec.spb.ru*)

Аннотация

В статье рассматриваются методы визуализации данных для анализа социальных сетей в интересах выявления и противодействия нежелательной и сомнительной информации. Приводятся ключевые принципы визуализации данных для социальных сетей и пример на основе визуальной аналитики нескольких социальных групп "ВКонтакте". Пример также содержит описание различных методов отрисовки и дает им оценку.

Ключевые слова: визуальная аналитика; визуализация социальных сетей; защита от информации; информационная безопасность; визуализация данных.

Введение

В настоящее время возрастает значимость и влияние социальных сетей на формирование человека как личности. В тоже время за счет широкого распространения в цифровом пространстве института репутации, у пользователей растет доверие к социальным сетям как к релевантному источнику информации в сравнении с традиционными каналами связи – телевидением, радио, журналами и т.д. Но помимо новостей, социальные сети стали площадкой распространения сомнительной и недостоверной информации, например, идей гомеопатии, отрицания холокоста, ВИЧ-диссидентства и т.д.

Оценивать масштаб распространения такой информации и тем более пресекать ее распространение крайне сложно [1]. Одним из способов обнаружения и анализа информационных потоков в социальных сетях является визуализация данных, которая позволяет в кратчайшие сроки в простом графическом виде представить сложные процессы [2-3].

В данной работе рассматриваются методы визуализации данных социальны сетей на примере графов и принципы визуализации социальных сетей в интересах выявления и противодействия нежелательной и сомнительной информации.

Визуализации данных социальных сетей

Методы визуальной аналитики должны соответствовать принципу Бена Шнейдермана [4]: иметь возможность обозревать данные целиком, иметь возможность масштабирования и возможность фильтрации. Ключевыми требованиями к визуализации можно считать также следующие принципы.

1) Возможность использования различных методов отрисовки. Для графов это могут быть радиальные, силовые или отрисовки на основе кластеризации.

2) Естественность методов человеко-машинного взаимодействия. Например, естественным способов визуализации 3D графов является дополненная реальность, а в 2D режиме при применении фильтров :можно использовать возможности сенсорных экранов.

Рассмотрим примеры выполнения этих принципов на основе визуальной аналитики нескольких социальных групп "ВКонтакте". Исходными данными является JSON-файл в котором используются следующие сведения: массив объектов который состоит из пользователей и групп; массив связей который состоит из отношений "репост" между пользователями и группами.

Самым простым способом представления является граф, который можно задать несколькими способами. Граф репостов между пользователями-группами, пользователями-пользователями и группами-группами изображен на рисунках 1 и 2 с помощью силового метода отрисовки графов.

[1] Аспирант, младший научный сотрудник.
[2] Аспирант, младший научный сотрудник.

Силовой метод отрисовки основывается на имитации физических процессов. Каждой вершине графа задается заряд, значение которого, в случае отрицательного значения, определяет силу, с которой вершины отталкиваются друг от друга, и, в случае положительного значения, притягиваются друг к другу. Каждому ребру графа при этом задается сила упругости на подобии с пружиной, которая определяет сопротивление силы отталкивания. Также ребрам задается значение длины пружины, которая определяет максимальной расстояние, на которое ребро может растянуться.

На рисунке 1 используются стандартные параметры для вершин и ребер графа. Очевидно, что силовой метод отрисовки упростил восприятие графа, и пространственно кластеризировал некоторые вершины. Тем не менее, в виду большой связности вершин, такой граф все еще является сложно-читаемым.

На рисунке 2 используются следующие параметры вершин:

1) Для вершин, имеющих одно ребро, устанавливается небольшое значение заряда, а для этого ребра устанавливается максимальный коэффициент упругости равный 1 (ребро не может растягиваться) и небольшая длина ребра.

2) Для вершин, которые имеют наибольшее количество ребер (статистически соответствуют “выбросам”), устанавливается большое значение заряда. Для ребер между такими вершинами устанавливается максимальный коэффициент упругости равный 1 (ребро не может растягиваться) и большая длина ребра.

3) Для всех остальных вершин устанавливается среднее значение заряда, а для всех остальных ребер устанавливается коэффициент упругости равный 0.1 (ребро может сильно растягиваться) и средняя длина ребра.



Рис. 1 Силовой метод отрисовки графа в котором используются стандартные параметры

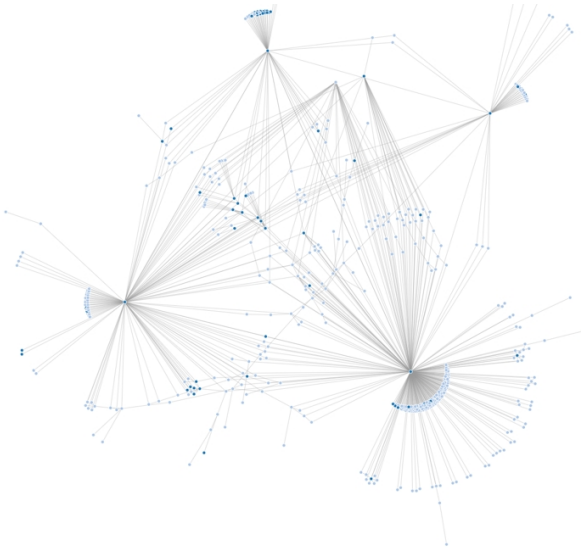


Рис. 2. Силовой метод отрисовки графа в котором для вершин и ребер используются индивидуальные параметры в зависимости от количества ребер у вершины

Очевидным преимуществом таких параметров силовой отрисовки является более выраженная пространственная кластеризация вершин. Вершины, которые имеют одно ребро и соответствуют пользователям которые репостят записи только из одной группы, группируются дугой вблизи вершины-группы. Вершины-пользователи группируются между общих вершин-групп. А вершины-группы располагаются достаточно далеко друг от друга, чтобы между ними можно было различить кластеры общих пользователей.

Еще одним способом отображения такого графа является отображение в 3D (рисунок 3). Отображение графа в 3D позволяет избежать наложения ребер друг на друга и еще больше пространственно кластеризирует вершины за счет использования третьего измерения.

Одним из требований является использование методов человеко-машинного взаимодействия. Рассмотрим применение данного требования на примере масштабирования изображения.

Если на 2D рисунках 1 и 2 масштабирование можно осуществлять путем простого увеличения изображения, то на 3D рисунке 3 необходимо имитировать камеру в 3D пространстве, которая имеет 3 двигательных и 3 вращательных степени свободы.

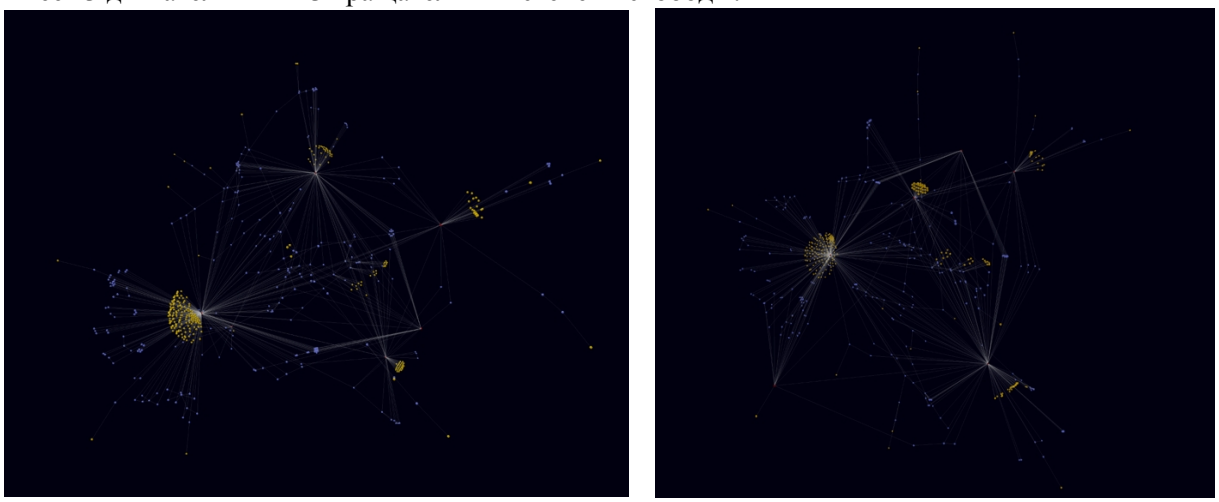


Рис. 3. Силовой метод отрисовки графа в 3D в двух ракурсах; исходные данные, использованные для построения графа, не отличаются от данных, использованных для графа на рисунке 2

Тем не менее, осуществлять навигацию в 3D пространстве с помощью клавиатуры и мыши неестественно для человека. Для этого можно использовать отображение графа в дополненной реальности (рисунок 4). Дополненная реальность делает процесс масштабирования и навигации более естественным: человек физически приближается, отдаляется и поворачивает голову, воспринимая граф как реальный физический объект, с которым можно взаимодействовать.

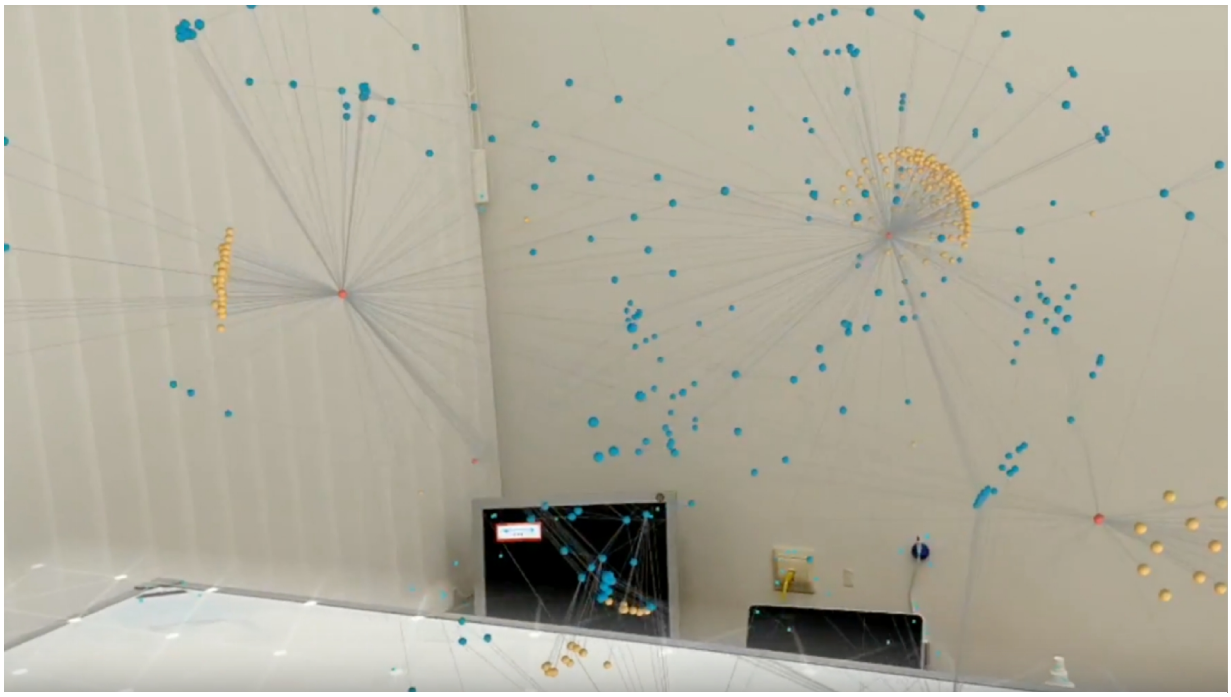


Рис. 4. Силовой метод 3D отрисовки графа в дополненной реальности; исходные данные, использованные для построения графа, не отличаются от данных, использованных для графа на рисунке 3 и 2

Таким образом, образуется сразу несколько моделей которые могут применяться при различных задачах. Например, для навигации в больших по размеру графах можно использовать графы в дополненной реальности, а для анализа локальных групп можно использовать 2D или 3D режим в зависимости от степени связанности вершин.

Заключение

В статье рассмотрены методы визуализации данных для анализа социальных сетей в интересах выявления и противодействия нежелательной и сомнительной информации. Приведены ключевые принципы визуализации данных для социальных сетей и пример на основе визуальной аналитики нескольких социальных групп “ВКонтакте”. Приведено описание различных методов отрисовки и дана им оценка.

Работа выполнена при финансовой поддержке Гранта РНФ (проект № 18-71-10094) в СПИИРАН.

ЛИТЕРАТУРА

1. Котенко И. В., Саенко И. Б., Чечулин А. А. Защита от нежелательной и вредоносной информации в глобальных информационных сетях // Информационно-психологическая и когнитивная безопасность. Коллективная монография / Под ред. И. Ф. Кефели, Р. М. Юсупова. СПб.: Изд-во «Аврора» 2017. 345 с. - С.207-229.
2. Комашинский Д.В., Котенко И.В., Чечулин А.А. Категорирование веб-сайтов для блокирования веб-страниц с неприемлемым содержимым // Системы высокой доступности, № 2, 2011. С.102-106.
3. Igor Kotenko, Andrey Chechulin, Andrey Shorov, Dmitry Komashinsky. Analysis and Evaluation of Web Pages Classification Techniques for Inappropriate Content Blocking. P. Perner (Ed.): 14th Industrial Conference on Data Mining (ICDM 2014), Lecture Notes in Artificial Intelligence (LNAI 8557), DOI 10.1007/978-3-319-08976-8. P.39–54. ISSN 0302-9743, ISBN 978-3-319-08975-1. (WoS, Scopus) (на английском).
4. Craft B., Cairns P. Beyond guidelines: what can we learn from the visual information seeking mantra? // Ninth International Conference on Information Visualisation, IEEE, 2005. – P. 110-118. (на английском)

M.V. Kolomeets, D.S. Levshun (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), St. Petersburg)

Data visualization for the identification and counteraction of undesirable information

Abstract. The paper examines methods of data visualization for the analysis of social networks in the interests of identifying and countering undesirable and questionable information. The key principles of data visualization for social networks and an example based on visual analysis of several social groups "VKontakte" are presented. The example also contains a description of the various rendering methods and gives them an estimation.

ЭМЕРДЖЕНТНЫЕ СВОЙСТВА КИБЕРФИЗИЧЕСКИХ СИСТЕМ

Д. С. Левшун[1], **Ю. Е. Бахтин**[2]
(*Санкт-Петербургский институт информатики и автоматизации
Российской академии наук (СПИИРАН),
199178, Россия, С.-Петербург, 14-я линия В.О., д. 39,
(812) 328-71-81, levshun@comsec.spb.ru, yurii.bakhtin@gmail.com*)

Аннотация

Современные киберфизические системы имеют многоуровневую архитектуру и представляют собой комбинацию различных подсистем. Объединение разнородных подсистем в рамках одной системы как расширяет их функциональные возможности, так и делает их подверженными атакам, одновременно происходящим на нескольких уровнях. Для противодействия подобным угрозам, необходимо обеспечить защищенность системы ещё на этапе её проектирования. Разработанная методика проектирования работает в рамках компонентного подхода, позволяя представить киберфизическую систему в виде взаимодействующих блоков. Объединение и взаимодействие блоков приводит к возникновению эмерджентных свойств, влияющих на функциональные возможности и нефункциональные ограничения как отдельных элементов, так и системы в целом. Данная работа направлена на классификацию эмерджентных свойств, а также улучшение процесса проверки киберфизических систем на корректность и реализуемость в рамках методики проектирования.

Ключевые слова: киберфизическая система, проектирование систем, эмерджентное свойство, микроконтроллер.

Введение

Современные киберфизические системы имеют ряд отличительных черт. Как правило, такие системы сложны, имеют многоуровневую архитектуру и представляют собой комбинацию различных подсистем. Например, современные решения в области обеспечения физической безопасности включают в себя как системы охранно-пожарной сигнализации, контроля доступа и видеонаблюдения, так и различные сервисы для повышения эффективности управления персоналом. А в основе современных решений в области кибернетической безопасности лежит построение полной модели инфраструктуры объекта защиты, а также распределенный сбор и анализ всей доступной информации об активах и событиях.

При этом, объединение разнородных подсистем в рамках единой системы имеет как достоинства, так и недостатки. С одной стороны, это приводит к расширению функциональных возможностей отдельных подсистем и системы в целом. С другой стороны, архитектурные особенности современных киберфизических систем делают их подверженными атакам, одновременно происходящим на нескольких уровнях, – от незаконного проникновения на объект защиты, до эксплуатации уязвимостей сетевого оборудования. При этом атаки могут быть разбиты на несколько шагов и растянуты во времени, а также нацелены на определенные подсистемы или устройства [1].

Возникновение эмерджентных свойств

Для противодействия подобным угрозам используются системы, реализованные в рамках комплексного подхода к обеспечению безопасности. Данный подход заключается в одновременной обработке разнородных событий кибернетического и физического уровней, а также в обеспечении защищенности киберфизической системы еще на этапе ее проектирования [2, 3]. При этом методики проектирования могут быть направлены на следующие элементы киберфизической системы:

- программное обеспечение;
- аппаратное обеспечение;
- отдельные устройства (программно-аппаратные элементы);

[1] Аспирант, младший научный сотрудник.
[2] Младший научный сотрудник.

- среду передачи данных;
- систему в целом.

В рамках предыдущих исследований был проведен анализ современных решений в данной области. По результатам проделанной работы, был сделан следующий вывод: "На данный момент не существует единого обобщенного подхода для решения задачи проектирования защищенных систем, содержащих устройства на основе микроконтроллеров, а существующие частные решения не лишены недостатков и нуждаются в доработке". Поэтому было принято решение о разработке нового подхода, ключевой особенностью которого является доработка уже существующих решений, а также их объединение с разработанными нами решениями в рамках единого цикла проектирования – Design Lifecycle of Secure Embedded Devices System (DLSEDS) [4]. DLSEDS представляет собой объединение (1) подхода к разработке безопасного программного обеспечения, (2) методики проектирования защищенных устройств на основе микроконтроллеров, а также разработанной (3) методики проектирования защищенных систем, основанной на результатах работы (1) и (2). При этом методика (3) выступает связующим звеном, формулирующим требования как к методике (1), так и к подходу (2). Кроме того, данная методика направлена на обеспечение защищенности среды передачи данных между элементами системы.

DLSEDS работает в рамках компонентного подхода, основная идея которого заключается в представлении киберфизических систем в виде совокупности взаимодействующих блоков (см. рис. 1). При этом каждый блок может состоять из других блоков, имеет функциональные требования и нефункциональные ограничения, выполнение которых необходимо для его применения в системе, а также функционал, который он способен системе предоставить. Задача методологии проектирования – предоставление наиболее рационального набора блоков с точки зрения предъявленных требований.

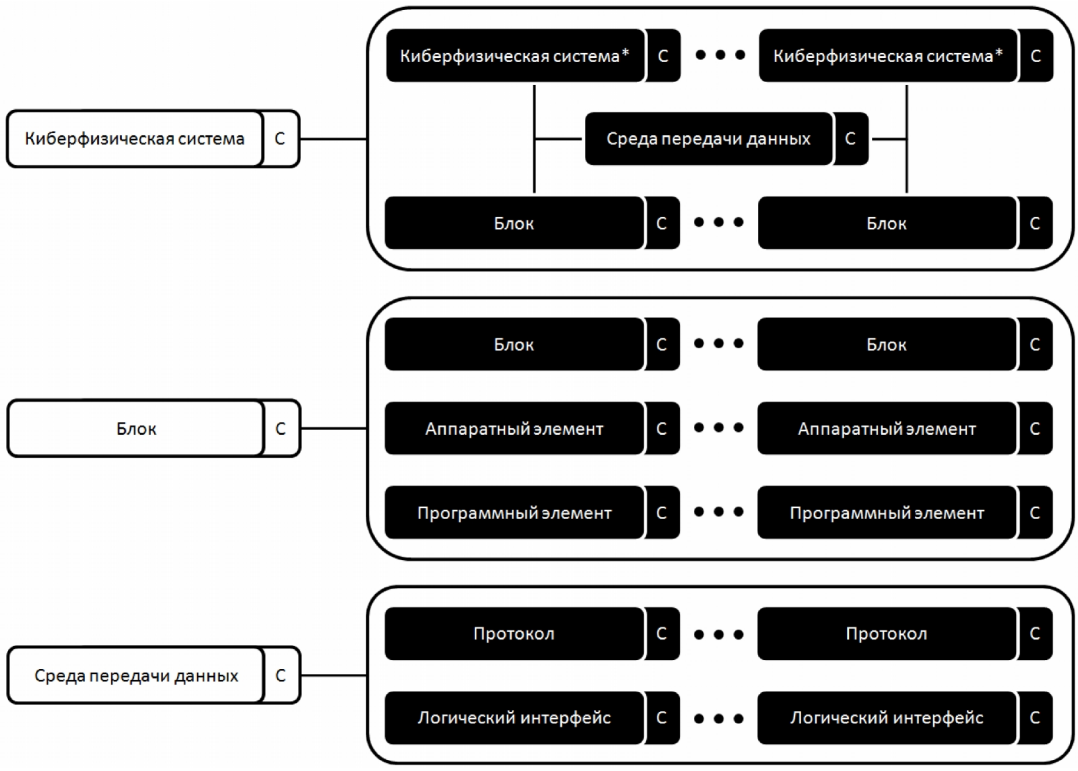


Рис. 1. Представление киберфизических систем в виде совокупности блоков

Объединение различных блоков и их взаимодействие в рамках единой системы приводит к возникновению эмерджентных свойств (свойств, проявляемых только в рамках системы и не соответствующих её элементам). Во время процесса проектирования в качестве эмерджентных

свойств рассматривается влияние объединения и/или взаимодействия отдельных блоков как при формировании других блоков, так и системы в целом. При этом эмерджентные свойства, повлиявшие на функциональные требования и нефункциональные ограничения отдельного блока, при формировании, например, другого блока на его основе уже не оказывают влияния, т.к. их эффект уже был учтён. Это позволяет при проектировании киберфизических систем сохранять определенный уровень абстракции во время формирования аналитических и имитационных моделей.

Классификация эмерджентных свойств

Влияние эмерджентных свойств при объединении и взаимодействии отдельных блоков киберфизических систем может быть классифицировано (1) по области влияния и (2) по результату влияния (см. рис. 2).



Рис. 2. Классификация эмерджентных свойств

По области влияния, эмерджентные свойства киберфизических систем могут быть разделены на (1) аппаратные, (2) программные, (3) программно-аппаратные и (4) системные. Рассмотрим каждое из них более подробно.

Аппаратные эмерджентные свойства проявляются при взаимодействии и объединении аппаратных элементов киберфизических систем. В качестве примера рассмотрим работу микроконтроллера Arduino Mega с несколькими RFID-считывателями с частотой 13.56 МГц (MFRC522). Полное считывание идентификатора метки с MFRC522 занимает 43 мс, при этом единовременно может работать только один RFID-считыватель, а сами считыватели не обладают внутренней памятью. Это означает, что RFID-метка, проведенная быстрее чем за 43 мс мимо двух RFID-считывателей (подсоединенных к одному микроконтроллеру), будет считана только одним из них.

Программные эмерджентные свойства проявляются при взаимодействии и объединении программных элементов киберфизических систем. В качестве примера рассмотрим построение безопасной и надежной среды передачи данных между одноплатными компьютерами Raspberry Pi и виртуальной машиной на базе операционной системы Ubuntu [5]. В качестве технологии построения среды передачи данных было выбрано решение на основе виртуальных частных сетей OpenVPN [6]. Безопасное и надежное VPN-соединение, в зависимости от выбранных на стороне клиента и сервера настроек, а также количества клиентов, оказывает различное влияние на пропускную способность используемой среды передачи данных, что и является проявлением программных эмерджентных свойств.

Программно-аппаратные эмерджентные свойства проявляются при объединении и взаимодействии программных, аппаратных и программно-аппаратных элементов киберфизических систем. В качестве примера рассмотрим организацию взаимодействия устройств на основе микроконтроллеров в рамках шины I2C [7]. Соединение по протоколу I2C между микроконтроллерами платформы Arduino организуется на основе аппаратного интерфейса TWI [8] и библиотеки Wire.h [9]. При этом скорость получения и обработки сообщений главным устройством напрямую зависит от количества подчиненных устройств, размера передаваемых сообщений (за один раз можно передать не более 32 байт данных), наличия подтверждения получения, взаимной аутентификации устройств и шифрования передаваемых данных.

Системные эмерджентные свойства проявляются на уровне проектируемой киберфизической системы при объединении и взаимодействии её элементов (аппаратных, программных и программно-аппаратных). В качестве примера рассмотрим систему распределенных вычислений, построенную на основе Hadoop [10] или Spark [11]. Очевидно, что взаимодействие нескольких вычислительных устройств для решения поставленной задачи подразумевает не только объединение их вычислительных ресурсов, но и выделение части этих ресурсов для разбиения задачи на более простые подзадачи, распределение подзадач и согласование работы устройств.

По результату влияния эмерджентные свойства киберфизических систем могут быть разделены на (1) положительные, (2) нейтральные и (3) отрицательные. При этом *нейтральные эмерджентные свойства* подразумевают собой отсутствие проявления эмерджентных свойств, однако они необходимы для полноты классификации. Рассмотрим положительные и отрицательные эмерджентные свойства более подробно.

Положительные эмерджентные свойства влекут за собой выполнение одного или нескольких функциональных требований и/или нефункциональных ограничений, необходимых для работы отдельных блоков киберфизических систем при их объединении и взаимодействии с другими блоками. В качестве примера рассмотрим подключение микроконтроллеров платформы Arduino к проводной сети Ethernet. Подключение Ethernet-чипа и интерфейса RJ45 к микроконтроллеру, как правило, связано с пайкой микроконтроллера для обеспечением необходимого тока и напряжения, а также взаимодействия через определенные PIN. В качестве альтернативы можно использовать уже готовые платы-расширения (например, Ethernet Shield), аппаратно и программно совместимые с микроконтроллерами платформы Arduino.

Отрицательные эмерджентные свойства влекут за собой или прекращение выполнения одного или нескольких функциональных требований и/или нефункциональных ограничений, необходимых для работы отдельных блоков киберфизических систем при их объединении и взаимодействии с другими блоками, или привносят дополнительные. В качестве примера дополнительных функциональных требований рассмотрим работы процессоров ATmega32U4 (прошивка устройств) и Atheros 9331 (OpenWrt [12], Linux) в рамках микроконтроллера Arduino Yún. Для полноценного вызова на уровне прошивки устройства linux-команд необходим специальный интерпретатор, роль которого, как правило, исполняет библиотека Bridge.h [13]. Наличие данной библиотеки и становится дополнительным функциональным требованием, которое также влияет на нефункциональные ограничения, связанные с допустимым объемом прошивки устройств и вычислительной сложностью решения.

При этом эмерджентные свойства могут иметь как область, так и результат влияния, позволяя выделять аппаратно-положительные, программно-отрицательные, системно-положительные и тому подобные свойства.

Заключение

Анализ эмерджентных свойств, возникающих в процессе объединения и взаимодействия различных элементов киберфизических систем, позволит значительно улучшить процесс проектирования киберфизических систем. Например, анализ влияния отрицательных эмерджентных свойств на функциональные требования и нефункциональные ограничения позволит выявлять возможные конфликты между элементами системы для последующего внесения информации о данных конфликтах в экспертную базу знаний. В свою очередь, анализ влияния

положительных эмерджентных свойств на функциональные требования и нефункциональные ограничения позволит выявлять наиболее рациональные наборы элементов, позволяющие сформировать итоговую киберфизическую систему.

Работа выполнена при поддержке бюджетной темы № АААА-А16-116033110102-5.

ЛИТЕРАТУРА

1. Kotenko I., Levshun D., Chechulin A. Event correlation in the integrated cyber-physical security system. Proceedings of the 2016 XIX IEEE International Conference on Soft Computing and Measurements (SCM-2016), IEEE, St. Petersburg, Russia, May 2016. P.484-486.
2. Desnitsky V., Levshun D., Chechulin A. and Kotenko I.. Design Technique for Secure Embedded Devices: Application for Creation of Integrated Cyber-Physical Security System. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), Vol.7, No.2, June, 2016. P.60-80.
3. Desnitsky V., Kotenko I., Chechulin A. Configuration-based approach to embedded device security // Lecture Notes in Computer Science, Springer-Verlag, 2012. Vol.7531. P.270-285.
4. Левшун Д.С., Чечулин А.А., Котенко И.В. Жизненный цикл разработки защищенных систем на основе встроенных устройств // Защита информации. Инсайд, № 4(76), 2017. С.53-59.
5. Официальный сайт операционной системы с открытым исходным кодом Ubuntu (основана на Debian GNU/Linux). – URL: https://www.ubuntu.com (дата обращения: 30.07.2018).
6. Официальный сайт технологии виртуальной частной сети с открытым исходным кодом OpenVPN. – URL: https://openvpn.net/ (дата обращения: 30.07.2018).
7. Dmitry Levshun, Andrey Chechulin, Igor Kotenko. A Technique for Design of Secure Data Transfer Environment: Application for I2C Protocol // The 1st IEEE International Conference on Industrial Cyber-Physical Systems (ICPS2018). Saint-Petersburg, Russia, May 15-18, 2018. P.789-794.
8. Официальная документация аппаратного интерфейса TWI для работы микроконтроллеров платформы Arduino в шине I2C. – URL: https://playground.arduino.cc/Main/WireLibraryDetailedReference#hardware (дата обращения: 30.07.2018).
9. Официальная документация для работы с шиной I2C на основе библиотеки Wire.h. – URL: https://www.arduino.cc/en/Reference/Wire (дата обращения: 30.07.2018).
10. Официальный сайт решения для надежных, масштабируемых и распределенных вычислений Hadoop. – URL: http://hadoop.apache.org/ (дата обращения: 30.07.2018).
11. Официальный сайт решения для надежных, масштабируемых и распределенных вычислений Spark. – URL: https://spark.apache.org/ (дата обращения: 30.07.2018).
12. Официальный сайт встроенной операционной системы с открытым исходным кодом OpenWrt (основана на ядре Linux). – URL: https://openwrt.org/ (дата обращения: 30.07.2018).
13. Официальная документация для работы процессоров ATmega32U4 и Atheros 9331 в рамках микроконтроллера Arduino Yún на основе библиотеки Bridge.h. – URL: https://www.arduino.cc/en/Reference/YunBridgeLibrary (дата обращения: 30.07.2018).

D.S. Levshun, Y.E. Bakhtin (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), St. Petersburg)

Emergency properties of cyber-physical systems

Abstract. Modern cyber-physical systems have a multi-level structure and contain different subsystems. Integration of various subsystems within a single system both expands their functionality and makes them vulnerable to attacks that occur simultaneously on several levels. To counteract such threats, it is necessary to ensure the security of the system even at the stage of its design. The developed design methodology works within the framework of the component approach, allowing to present the cyber-physical system in the form of interacting blocks. The combination and interaction of blocks leads to the emergent properties that affect the functional requirements and non-functional limitations of both individual elements and the system. This paper is devoted to the classification of emergent properties, as well as the improvement on its basis of the model check process of the system in terms of its correctness and realizability.

МЕТОДЫ ПРОФИЛИРОВАНИЯ И ВЫЯВЛЕНИЯ АНОМАЛИЙ ДЛЯ ОБНАРУЖЕНИЯ ВРЕДОНОСНОЙ АКТИВНОСТИ

Н. А. Комашинский
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук, С.-Петербург, 14 линия В.О., д. 39, nckkm@ya.ru*)

Аннотация

В работе рассматриваются индустриальные решения, основанные на математической статистике и анализе поведения пользователей по данным о работе с информационными и вычислительными ресурсами компьютерной системы, базирующиеся на двух основных технологиях: технология анализа поведения пользователей и систем (User and Entity Behavior Analytics, UEBA) и технология управления информацией и событиями безопасности (Security Information and Event Management, SIEM). Применение SIEM-решений обособленно слабо справляется со всеми проблемами безопасности – для их работы необходимо разрабатывать и затем поддерживать работу правил корреляции, а также всегда знать, что именно искать. Таким образом, с помощью решений, основанных на машинном обучении и ретроспективном анализе решения UEBA, можно обнаруживать такие инциденты, которые классические системы пропускают.

Ключевые слова: Аномалия, поведенческий анализ, SIEM, математическая статистика, корреляция.

Введение

Одним из новых, но достаточно эффективных методов раннего обнаружения компьютерных инцидентов стал анализ поведения пользователей и выявление аномалий в их поведении. Подход подразумевает, что необычная активность, не соответствующая стандартному профилю, может более точно указывать на потенциальные инциденты информационной безопасности. По классификации Gartner, этот класс решений называется User and Entity Behavior Analytics (UEBA).

Решения этого класса анализируют образцы поведения пользователей, применяя специализированные алгоритмы и статистический анализ для обнаружения значительных аномалий в поведении, указывающих на потенциальные угрозы. Таким образом, UEBA призваны заблаговременно обнаружить и предупредить об отклонениях в поведении.

Методов обнаружения аномалий [1] существует огромное количество, а также их модификаций и разновидностей (рис. 1).

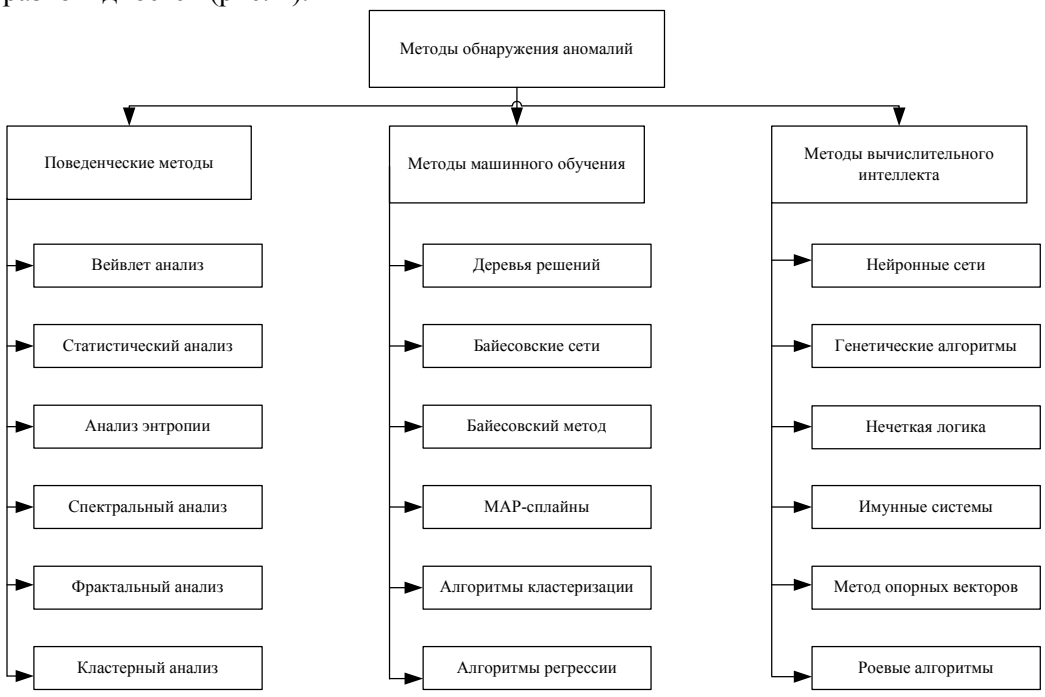


Рис. 1. Методы обнаружения аномалий

Основной целью построения и функционирования SIEM-систем является повышение уровня информационной безопасности в защищаемой компьютерной системе за счет обеспечения в режиме, близком к реальному времени, возможности манипулировать информацией о безопасности и осуществлять управление инцидентами и событиями безопасности, предупреждая возможность перехода защищаемой компьютерной системы в критическое с точки зрения информационной безопасности состояние [2-6].

Тогда как традиционная технология управления информацией и событиями безопасности (SIEM) поддерживает мониторинг и анализ действий на основе сигнатур с применением пользовательского контекста, технологии UEBA расширяют SIEM, осуществляя более эффективный мониторинг и анализ исключений (аномалий). Многие решения, основанные на технологии UEBA, поддерживают интеграцию с SIEM-системами для более всестороннего анализа. Основанные на машинном обучении и ретроспективном анализе решения UEBA-системы позволяют увидеть, что делал пользователь, например, полгода назад и как изменилась его активность в настоящее время. Наряду с применением методов математической статистики это называется профилированием; с его помощью можно обнаружить такие инциденты, которые классические системы пропускают.

В докладе предлагается метод диагностики аномалий, основанный на разделении многомерного пространства с набором измерений сетевого трафика, на непересекающиеся подпространства, соответствующие нормальным и аномальным сетевым условиям. Данное разделение может быть эффективно выполнено с помощью анализа основных компонентов.

Основная часть

Постановка задачи. Проблема диагностики аномалии в сетевом потоке может быть разделена на три этапа: (1) обнаружение, (2) идентификация и (3) количественное определение. Проблема обнаружения состоит в определении моментов времени, где присутствует нестандартное поведение сети. Эффективный алгоритм решения проблемы обнаружения должен обладать высокой вероятностью обнаружения и низкой вероятностью ложной тревоги. Задача идентификации состоит в том, чтобы выбрать тип истинной аномалии из множества возможных экземпляров аномалий. Наконец, количественная оценка представляет собой проблему оценки количества дополнительных или отсутствующих байтов в базовых потоках трафика. Количественная оценка важна, поскольку она дает представление о важности аномалии. Для успешной диагностики необходимо уметь определять время аномалии, идентифицировать основной поток трафика и количественно определять размер аномалии внутри этого потока.

Диагностика аномалий. Обозначим множество нормальных и аномальных состояний про-

странства S и соответственно. Обнаружение аномалий в трафике зависит от разделения канала y при любом временном значении на нормальные и аномальные компоненты. Обозначим их как смоделированные и остаточные части y. Основная идея на этапе обнаружения на

основе подпространства состоит в том, что после того, как S и были построены, это разделение может быть эффективно выполнено путем формирования проекции трафика на эти два подпространства. Предполагаемый набор измерений связей в заданный момент времени y:

$$y = \hat{y} + \tilde{y}$$

где $\tilde{y}$ соответствует смоделированному, а остаточному трафику. Формируем $\tilde{y}$, проеци-

руя y на S и формируем проецируя y на . Упорядочим множество главных компонент, соответствующих нормальному подпространству ($v_1, v_2, \ldots, v_r$), как столбцы матрицы P

размера $m \times r$, где r обозначает число нормальных осей. Затем можно записать $\tilde{y}$ и как:

$$\tilde{y} = PP_y^T = C_y \quad \text{и} \quad \tilde{y} = (I - PP^T)_y = \tilde{C}_y$$

где матрица $C = PP^T$ представляет собой линейный оператор, выполняющий проекцию на нормальное подпространство S, а $\tilde{C}$ проецирует на подпространство аномалий $\tilde{S}$.

Полезной статистикой для выявления аномальных изменений в является квадратичная ошибка прогнозирования (*SPE*):

$$SPE \equiv \|\tilde{y}\|^2 = \|\tilde{C}_y\|^2$$

и можно считать, что сетевой трафик является нормальным, если: $SPE \leq \delta_\alpha^2$, где δ_α^2 обозначает порог для *SPE* на уровне достоверности $1 - \alpha$. Статистический тест для остаточного вектора, известного как Q-статистика был разработан Джексоном и Мудхолкаром [7].

$$\delta_\alpha^2 = \varphi_1 \left[\frac{C_\alpha \sqrt{2\varphi_2 h_0^2}}{\varphi_1} + 1 + \frac{\varphi_2 h_0 (h_0 - 1)}{\varphi_1^2} \right]^{\frac{1}{h_0}},$$

где $h_0 = 1 - \frac{2\varphi_1 \varphi_3}{3\varphi_2^2}$ и $\varphi_i = \sum_{j=r+1}^{m} \lambda_j^i$ для $i = 1, 2, 3$ и где λ_j – дисперсия, полученная путем проецирования данных по j-му главному компоненту $\left(\|Y_{v_j}\|^2\right)$ и C_α это $1 - \alpha$ перцентиль в стандартном нормальном распределении. Результат Джексона и Мудхолкара выполняется независимо от того, сколько основных компонентов сохраняется в нормальном подпространстве.

Стоит отметить, что доверительный предел $1 - \alpha$ соответствует ложному оповещению коэффициента α, если выполнены предположения, при которых этот результат получен.

Предел доверия для Q-статистики выведен в предположении, что вектор образца y следует за многомерным распределением Гаусса. Однако в [8] отмечается, что Q-статистика мало меняется, даже когда основное распределение исходных данных существенно отличается от Гауссова. Предполагая, что нормальный трафик в наборах данных достаточно хорошо описывается как многомерный Гауссовский, не изучены данные о нарушениях этого предположения. Однако Q-статистика дает превосходные результаты на практике [8].

Важным свойством этого подхода является то, что он не зависит от среднего объема трафика в сети. Таким образом, один и тот же тест можно применить к сетям разного размера и уровням использования.

Такой подход к проверке сосредоточен на ответах на два вопроса: (1) насколько хорошо метод может диагностировать фактические аномалии, наблюдаемые в реальных данных и (2) как влияет время и место аномалии на эффективность метода.

Для решения первого вопроса используется анализ временных рядов по данным потока трафика, сначала изолируется набор «истинных» аномалий. Это позволяет затем количественно оценить метод подпространства. В частности, это позволяет измерять как вероятность обнаружения, так и вероятность ложной тревоги.

Для ответа на второй вопрос, вводятся аномалии разного размера в потоках трафика и применяется процедура для диагностики известных аномалий из данных соединения. Выполняется это многократно для каждого временного интервала и для каждой аномалии, чтобы составить полную картину того, насколько эффективность диагностики зависит от времени и места аномалии. В каждом случае оценивается эффективность каждого шага в процедуре диагностик. Успех обнаружения измеряется с помощью двух метрик: скоростью обнаружения и ложным сигналом тревоги. Скорость обнаружения - это доля обнаруженных истинных аномалий. Уровень ложной тревоги - это доля нормальных измерений, которые вызывают ошибочное обнаружение. Успех идентификации фиксируется в скорости идентификации, которая представляет собой долю обнаруженных аномалий, которые были правильно идентифицированы. Наконец, успех количественной оценки измеряется путем вычисления средней абсолютной относительной погрешности между оценкой и истинным размером всех выявленных аномалий объема.

Чтобы идентифицировать набор «истинных» аномалий в наших данных, мы ищатся необычные отклонения от среднего значения в каждом потоке трафика. Существует два общих класса методов для обнаружения таких изменений в нестационарных временных рядах. Первый класс методов идентифицирует аномалии в режиме реального времени, основанные на грубых отклонениях от прогнозируемого поведения. Простыми примерами такой стратегии являются алгоритмы прогнозирования скользящего среднего и модель Холь-та-Уинтерса [9]; более сложными примерами являются модели прогнозирования Бокса-Дженкинса [9]. Второй класс методов обрабатывает данные в пакетах и основывается на методах анализа сигналов, таких как схема вейвлет-анализа [10]. Такие схемы моделируют временные интервалы, выделяя низкочастотные компоненты; аномалии помечены в те моменты времени, которые значительно отличаются от моделируемого поведения среднего значения. Следует отметить, что ни одна схема не идеальна для выделения и количественного определения пиков во временных сериях и что некоторые схемы чувствительны к набору параметров конфигурации и предположений моделирования. Таким образом, для получения наборов истинных аномалий использовался метод кандидата из каждого класса методов.

Заключение

Предложенные методы обнаружения аномалий в большом объеме трафика можно использовать в системах поведенческого анализа. Эти схемы работают на основе измеряемого по сетевому каналу трафику, независимо от трафика по другим сетевым каналам. Таким образом, эти методы используют временные шаблоны в одном временном ряде трафика для выявления аномалий. Предложенные методы используют свойства корреляции между сетевыми соединениями для обнаружения аномалий в масштабах сети. Еще одна отличительная особенность такого подхода заключается в том, что он не требует подробных предложений моделирования относительно нормального поведения трафика. Предыдущие методы в основном основывались на моделях временных рядов для приближения к нормальному трафику [9] или полагались на построение моделей данных в течение периода обучения [11]. Например, широко используемый алгоритм Холь-та-Уинтерса создает модель для нормального поведения сетевого потока на основе параметров и априорного знания периодической структуры трафика. Разработанная схема не полагается на какие-либо параметры, и нормальное поведение трафика фиксируется непосредственно в данных, путем извлечения общих функций по всем сетевым взаимодействиям.

Работа проводилась при частичной финансовой поддержке РФФИ (проекты 16-29-09482, 18-07-01369 и 18-07-01488) и бюджетной темы № АААА-А16-116033110102-5.

ЛИТЕРАТУРА

1. Котенко И.В., Саенко И.Б., Агеев С.А., Копчак Я.М. Обнаружение аномального трафика в сетях Интернета вещей на основе нечеткого логического вывода // XVIII Международная конференция по мягким вычислениям и измерениям (SCM'2015). Сборник докладов.Том 1. СПб.: Издательство СПбГЭТУ «ЛЭТИ». 2015. С.9–14.
2. Котенко И.В., Саенко И.Б. SIEM-системы для мониторинга и управления инцидентами // Транспортная безопасность и технологии. 2017. № 04(51). С. 96-97.
3. Котенко И.В., Саенко И.Б. Архитектура системы интеллектуальных сервисов защиты информации в критически важных инфраструктурах // Труды СПИИРАН. № 1 (24). 2013. С.21-40.
4. Котенко Д.И., Котенко И.В., Саенко И.Б. Методы и средства моделирования распределенных атак в больших компьютерных сетях: состояние проблемы // Труды СПИИРАН. № 3 (22). 2012. С.5-30.
5. Kotenko I., Konovalov A., Shorov A. Agent-based simulation of cooperative defence against botnets // Concurrency Computation Practice and Experience. 2012. Vol.24. № 6. P.573-588.
6. Кожевникова И. С. Анализ методов обнаружения аномалий для обнаружения сканирования портов // Молодой ученый. — 2017. — №14. — С. 31-34.
7. Jackson J. E. and Mudholkar G. S. Control Procedures for Residuals Associated with Principal Component Analysis. Technometrics, pages 341–349, 1979.
8. D. R. Jensen and H. Solomon. A Gaussian Approximation to the Distribution of a Definite Quadratic Form. J. of the American Statistical Association, pages 898–902, 1972.
9. B. Krishnamurthy, S. Sen, Y. Zhang, and Y. Chen. Sketch-based Change Detection: Methods, Evaluation, and Applications. In Internet Measurement Conference, 2003.
10. Barford P., Kline J., Plonka D., and Ron A. A Signal Analysis of Network Traffic Anomalies. In Internet Measurement Workshop, 2002.
11. Ye N. A Markov Chain Model of Temporal Behavior for Anomaly Detection. In Workshop on Information Assurance and Security, 2000.

N.A. Komashinsky (St. Petersburg Institute for Informatics and Automation of the Russian Academy)
Methods of profiling and anomaly detection for detecting malicious activity

The paper deals with industrial solutions based on mathematical statistics and analysis of user's behavior based on data on work with information and computing resources of a computer system based on two main technologies: User and Entity Behavior Analytics (UEBA) and control technology information and security events (SIEM). The use of SIEM-solutions alone weakly copes with all security problems - for their work it is necessary to develop and then maintain the work of correlation rules, and also always know what exactly to look for. Thus, using solutions based on machine learning and retrospective analysis of the UEBA solution, it is possible to detect incidents that classical systems miss.

СИСТЕМА ПОКАЗАТЕЛЕЙ И КРИТЕРИЕВ ОЦЕНКИ КАЧЕСТВА ПОЛИТИК РАЗГРАНИЧЕНИЯ ДОСТУПА В ОБЛАЧНЫХ ИНФРАСТРУКТУРАХ КРИТИЧЕСКИ ВАЖНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

И. Б. Саенко[1]
(*Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский институт информатики и автоматизации»
Российской академии наук (СПИИРАН),
199178, С.-Петербург, 14-я линия В.О., д. 39,
(812) 328-71-81, ibsaen@comsec.spb.ru*)

В. И. Комашинский[2]
(*Федеральное государственное бюджетное учреждение науки «Институт проблем транспорта Российской академии наук им. Н.С. Соломенко» (ИПТ РАН),
199178, С.-Петербург, 12-я линия В.О., 13,
+7(812)-323-29-54, info@iptran.ru*)

Аннотация

В докладе рассматриваются вопросы построения системы показателей и критериев оценки качества политик разграничения доступа, применяемых в облачных инфраструктурах критически важных информационных систем. Приведена характеристика современных моделей разграничения доступа для облачных инфраструктур. Обоснованы требования и дано формальное описание показателей и критериев качества политик разграничения доступа.

Ключевые слова: политика разграничения доступа, облачная инфраструктура, критически важная информационная система.

Введение

Повышение роли и стремительное распространение облачных инфраструктур в критически важных информационных системах, повышение совокупной стоимости активов устройств, программного обеспечения и критически важных данных таких систем, а также увеличение числа атак на них определяют актуальность задач разграничения доступа к информации в облачных инфраструктурах таких систем, а также обнаружения и разрешения конфликтов в используемых ими политиках разграничения доступа. К числу ключевых тенденций развития облачных инфраструктур относят дальнейшее увеличение ее возможностей по обеспечению надежности, безопасности, производительности, управляемости и масштабируемости. Таким образом, вопросы обеспечения безопасности, а разграничение доступа является составной частью безопасности, стоят на одном из первых мест.

Одной из важнейших задач, решаемых при разработке системы разграничения доступа к облачным инфраструктурам, является разработка и обоснование системы показателей и критериев оценки качества политик разграничения доступа для различных моделей управления доступом, применяемых в облачных инфраструктурах критически важных информационных систем. Показатели качества политик разграничения доступа должны показывать, насколько действующие политики отвечают предъявляемым требованиям, и позволять сформировать различные критерии синтеза (оптимизации) и реконфигурации этих политик. Рассмотрение ключевых решений по построению системы показателей и критериев оценки качества политик разграничения доступа в облачных инфраструктурах критически важных информационных систем является целью настоящего доклада.

Обоснование показателей и критериев качества политик разграничения доступа

Характеристика современных моделей разграничения доступа для облачных инфраструктур. Облачные инфраструктуры представляют собой сравнительно новый вид компью-

[1] Доктор технических наук, профессор, ведущий научный сотрудник.
[2] Доктор технических наук, доцент, заместитель директора.

терных инфраструктур, которые притягивают к себе повышенное внимание в области информационных технологий и обладают повышенным интересом для потребителей информационных услуг и ресурсов. В то же время многие исследователи отмечают, что более широкому распространенному этой технологии препятствуют проблемы безопасности [1-6]. Недостатком большинства существующих облачных инфраструктур является отсутствие возможности гибкого управления со стороны пользователей доступом к своим данным, что вызвано универсальностью решений по контролю доступа, принимаемых поставщиками облачных услуг [1]. Многие исследователи отмечают, что неоднородность и большое разнообразие ресурсной среды облачного хранилища требуют всестороннего и детально проработанного механизма управления доступом, чтобы обеспечить динамические, постоянно расширяемые и хорошо настраиваемые требования по защите информации пользователей [2]. Однако существующие механизмы безопасности, обеспечиваемые поставщиками облачной инфраструктуры, не удовлетворяют этим требованиям [3]. Кроме того, проблем безопасности информации в облачных инфраструктурах обостряются, если используются открытые веб-сервисы [4]. Все это настоятельно требует проработки вопросов совершенствования политик разграничения доступа и моделей, лежащих в их основе [5, 6].

Задача обеспечения требуемого разграничения доступа к ресурсам информационной системы и защиты этих ресурсов от несанкционированного доступа, особенно со стороны инсайдеров, несмотря на имеющиеся в этой области научные и практические результаты, по-прежнему остается актуальной практически для всех разновидностей критически важных информационных систем независимо от их предназначения. В еще большей степени ее актуальность возрастает в случае применения облачных инфраструктур для объединения услуг и ресурсов, что влечет за собой объединения и интеграцию частных политик разграничения доступа. Значительное возрастание числа пользователей и прикладных программ, требующих доступа к ресурсам облачной инфраструктуры, резкое возрастание номенклатуры и объема защищаемых объектов доступа и разновидностей применяемых к ним полномочий доступа, высокая динамика изменения состава ресурсов облачного хранилища в процессе функционирования критически важной информационной системы, разнородность средств и механизмов контроля и разграничения доступа к информации, которыми обладают автоматизированные системы, охваченные облачной инфраструктурой, – все это приводит к тому, что системы защиты и административный персонал, оснащенные арсеналом традиционных средств разграничения доступа, не могут своевременно выявлять уязвимости в политиках и схемах разграничения доступа, обнаруживать инциденты безопасности и адекватно реагировать на попытки осуществления несанкционированного доступа. Во многом это обусловлено недостаточной эффективностью использования традиционных моделей контроля доступа (access control model) к информационным ресурсам, таких как дискреционная (Discretionary access control, DAC), мандатная (Mandatory access control, MAC), ролевая (Role-based access control, RBAC) и механизмов их реализации, реализованных в операционных системах, системах управления базами данных, геоинформационных системах, системах управления предприятиями и других компонентах программного обеспечения, на основе которых в настоящее время возможно формирование облачной инфраструктуры. По этой причине представляется целесообразным и достаточно интересным в научном плане подход, при котором для облачных инфраструктур вводятся в рассмотрение более гибкие модели контроля доступа, содержащие элементы искусственного интеллекта, например, такие как атрибутивно-ориентированная (attribute-based access control, ABAC), организационно-ориентированная (Organization-based access control, OrBAC), онтолого-ориентированная (Ontology-based access control, OBAC) и другие. Данные модели контроля доступа являются сравнительно новыми, их известные случаи реализации носят исследовательский и/или экспериментальный характер.

Анализируя модели контроля доступа к информации, которые применяются в облачных инфраструктурах, можно сделать вывод, что наибольшей популярностью обладает модель RBAC, в дополнение к которой предлагается использовать некоторые более перспективные модели. В [5], кроме RBAC, анализируются и предлагаются к использованию модель ABAC и модель «множественной аренды» (multi-tenancy), являющаяся разновидностью OrBAC. В [3] предлагается использовать основанную на RBAC модель Amazon Web Services (AWS), позволяющую

расширить возможности политик RBAC и облегчить интеграцию службы разграничения доступа в корпоративные приложения. В [2] предлагается гибкая междоменная модель разграничения доступа, основанная на механизмах преобразования ролей. В [6] среди анализируемых моделей доступа присутствуют MAC и DAC, но сразу же отмечается их низкая эффективность в облачных инфраструктурах. В качестве наиболее приемлемых предлагается использование моделей RBAC и ABAC. В [4] предлагается так называемая модель, основанная на политиках (Policy-based Access Control), которая является разновидностью модели OBAC. В [1, 3] предлагается реализовать новую парадигму услуг, связанную с разграничением доступа, так называемый «контроль доступа как сервис» (Access Control as a Service) с использованием применяемых в облачных инфраструктурах моделей доступа.

Вопросы оценки, формирования и оптимизации политик разграничения доступа в научной литературе в наибольшей степени проработаны для модели RBAC, т.к. она допускает строгую формализацию с выделением переменных, целевой функции и ограничений. Это позволило выделить задачу оптимизации политики разграничения доступа на основе RBAC в отдельный класс задач интеллектуального анализа данных (Data Mining), получившего название Role Mining Problem (RMP) [7]. Разработаны различные варианты постановки этой задачи и предложен ряд методов и алгоритмов их решения [8].

В качестве более общего подхода для решения задач RMP [9, 10] предлагается использовать генетические алгоритмы. В этих работах показано, что генетические алгоритмы, как и другие алгоритмы биоинспирированной оптимизации, следует рассматривать как достаточно эффективные средства решения задач оптимизации политик разграничения доступа.

В ряде работ частично рассматриваются отдельные вопросы, связанные с задачей реконфигурирования схемы RBAC. Такими работами являются [11-13]. Так, в [11] предложен подход, согласно которому для реконфигурации схемы RMP используются access history logs. Однако при этом не ставится вопрос о минимизации административных затрат. В [12] показана возможность реконфигурации схемы RMP с помощью так называемых «минусовых связей». Однако учет таких связей в общем случае требует создания специального программного обеспечения. В [13] предложены постановки RMP, в которых, однако, имеются не свойственные для облачных инфраструктур дополнительные ограничения.

В [14] предложен подход для верификации политик разграничения доступ, основанный на адаптации различных ситуаций с помощью моделирования уступок, которые пользователи делают, чтобы достигнуть разрешения возможных конфликтов. Этот подход был рассмотрен применительно к социальным системам, однако он представляется применимым для облачных инфраструктур и планируется к использованию в настоящем проекте. В [15] для верификации политик разграничения доступа предлагается использование эвристик, которые понижают сложность механизма разграничения доступа.

Из характеристики современных моделей разграничения доступа для облачных инфраструктур можно сделать следующие выводы. Во-первых, проблема разграничения доступа в облачных инфраструктурах вызвана спецификой свойственных им угроз безопасности и является достаточно актуальной. Эта проблема обостряется в критически важных информационных системах. Во-вторых, решение этой проблемы связывается с использованием наряду с традиционной моделью контроля доступа RBAC новых перспективных моделей ABAC, OrBAC, OBAC и других, построенных на их основе.

Требования к показателям и критериям качества политик разграничения доступа. Оценка качества политик разграничения доступа в облачной инфраструктуре необходима для выявления необходимости принятия контрмер по противодействию атакам на систему разграничения доступа и определения условий начала выполнения реконфигурация политик разграничения доступа. Согласно этому подходу, оценка сводится к сопоставлению требуемой и результирующей схем разграничения доступа.

Требуемая схема разграничения доступа задается лицом, принимающим решения (ЛПР). Результирующая схема образуется на основании правил, свойственных выбранной модели контроля доступа, применяемых в сформированной конфигурации политики разграничения доступа. Мера близости требуемой и результирующей схем является одним из показателей качества разработанной политики разграничения доступа.

Другими показателями могут случить различные структурные параметры, например, количество ролей (для модели RBAC), общее количество связей и т.д. В настоящий момент показатели качества политики разграничения доступа в достаточной степени разработаны только для модели RBAC. На их основе сформированы различные варианты оптимизационной задачи, известной как «проблема извлечения ролей» (Role Mining Problem, RMP). Для моделей ABAC, OrBAC, OBAC и прочих этот вопрос является пока не исследованным.

Формальное описание показателей и критериев политик разграничения доступа. Введем следующие обозначения:

1) $U = \{u_i\},\ i = 1,..,I$ – множество объектов доступа;

2) $R = \{r_j\},\ j = 1,..,J$ – множество субъектов доступа;

3) $V = \{v_k\}$ – множество типов полномочий, которые субъекты доступа имеют по отношению к объектам доступа;

4) $H_{\text{СРД}}$ – множество средств, входящих в состав системы разграничения доступа;

5) $S^{\text{треб}} \subseteq U \times R \times V$ – требуемая схема разграничения доступа субъектов к объектам доступа.

Структура и параметры организации $H_{\text{СРД}}$ определяют вид реальной схемы разграничения доступа $S^{\text{реал}} \subseteq U \times R \times V$, которая задается функционалом вида

$$S^{\text{реал}} = F\left(U, R, V, H_{\text{СРД}}\right). \quad (1)$$

Введем обозначения: $\Delta S^{\text{треб}} = S^{\text{треб}} \setminus S^{\text{реал}}$ и $\Delta S^{\text{реал}} = S^{\text{реал}} \setminus S^{\text{треб}}$.

Если $S^{\text{реал}}$ совпадает с $S^{\text{треб}}$, то выполняются условия

$$\Delta S^{\text{треб}} = 0, \quad (2)$$

$$\Delta S^{\text{реал}} = 0. \quad (3)$$

В этом случае при построении системы разграничения доступа полностью удовлетворяются требования по обеспечению безопасности информации. Однако на практике, как правило, либо не выполняется (2), либо не выполняется (3), либо не выполняются (2) и (3) вместе.

Невыполнение условия (2) означает, что требуемая схема $S^{\text{треб}}$ содержит элементы, которые отсутствуют в реальной схеме $S^{\text{реал}}$. В результате имеются некоторые требования, которые не обеспечены реальной схемой. Это означает, что имеет место нарушение свойства конфиденциальности информации. Невыполнение условия (3), напротив, означает, что имеются избыточные ограничения на доступ пользователей к информации, не подтвержденные требованиями, так как реальная схема $S^{\text{реал}}$ содержит элементы, которые отсутствуют в требуемой схеме $S^{\text{треб}}$. По сути это означает, что имеет место нарушение свойства доступности информации. Так как свойства конфиденциальности и доступности являются частными свойствами безопасности информации как свойства более высокого уровня, нарушение любого из указанных частных свойств означает нарушение безопасности информации в целом. Поэтому для построения критериев качества политик разграничения доступа следует использовать показатели, характеризующие выполнение со стороны системы разграничения доступа требований по конфиденциальности и доступности информации.

Показатель конфиденциальности $W^{\text{конф}}$ и показатель доступности $W^{\text{дост}}$ задаются соответствующими функционалами

$$W^{\text{конф}} = W^{\text{конф}}\left(\Delta S^{\text{треб}}\right), \quad (4)$$

$$W^{\text{дост}} = W^{\text{дост}}\left(\Delta S^{\text{реал}}\right). \quad (5)$$

Показатель безопасности информации в облачной инфраструктуре $W^{\text{бо}}$ в общем случае задается как

$$W^{\text{бо}} = W^{\text{бо}}\left(\Delta S^{\text{треб}}, \Delta S^{\text{реал}}\right). \quad (6)$$

В частных случаях, когда $W^{\text{конф}}$ и $W^{\text{дост}}$ являются независимыми, вместо (6) возможно использование выражения

$$W^{\text{бо}} = W^{\text{бо}}\left(W^{\text{конф}}, W^{\text{дост}}\right). \quad (7)$$

Вид функционалов (4), (5), (6) и (7) определяется конкретным составом и структурой системы разграничения доступа и используемыми для ее построения моделями разграничения доступа.

Если имеет место использование в явном виде (4) и (5), то критерии оценки качества политик разграничения доступа формируются в следующих вариантах:

$$W^{\text{конф}} \geq W^{\text{конф}}_{\text{треб}}, W^{\text{дост}} \Rightarrow \max. \quad (8)$$

$$W^{\text{конф}} \geq W^{\text{конф}}_{\text{треб}}, W^{\text{дост}} \Rightarrow \max. \quad (9)$$

$$W^{\text{конф}} \geq W^{\text{конф}}_{\text{треб}}, W^{\text{дост}} \geq W^{\text{дост}}_{\text{треб}}. \quad (10)$$

где $W^{\text{конф}}_{\text{треб}}$ и $W^{\text{дост}}_{\text{треб}}$ – требуемые значения показателей конфиденциальности и доступности, соответственно.

Если имеет место использование (6) или (7), то критерий имеет один из следующих видов:

$$W^{\text{бо}} \geq W^{\text{бо}}_{\text{треб}}, . \quad (11)$$

$$W^{\text{бо}} \Rightarrow \max. \quad (12)$$

где $W^{\text{бо}}_{\text{треб}}$ – требования, предъявляемые к системе разграничения доступа по обеспечению безопасности информации.

Следует отметить, что предложенные показатели и критерии носят универсальный характер. Они не зависят от используемых моделей разграничения доступа. Следовательно, они могут быть непосредственно применимы для разработки методов и методик оптимизации политик разграничения доступа, применяемых в облачных инфраструктурах критически важных информационных систем.

Заключение

Разработка системы показателей и критериев оценки качества политик разграничения доступа являются одной из важнейших задач обеспечения информационной безопасности в облачных инфраструктурах критически важных информационных систем. Решение этой задачи должно учитывать не только традиционные модели контроля доступа, но и перспективные модели. Оценка качества политик разграничения доступа сводится к сопоставлению требуемой и результирующей схем доступа. Предлагаемые формальные описания показателей и критериев политик разграничения доступа являются универсальными и могут быть применены как к традиционным, так и к перспективным моделям разграничения доступа, используемым в облачных инфраструктурах критически важных информационных систем.

Работа проводилась при частичной финансовой поддержке РФФИ (проекты 16-29-09482, 18-07-01369 и 18-07-01488) и бюджетной темы № АААА-А16-116033110102-5.

ЛИТЕРАТУРА

1. Саенко И.Б., Бирюков М.А., Ясинский С.А., Грязев А.Н. Реализация критериев безопасности при построении единой системы разграничения доступа к информационным ресурсам в облачных инфраструктурах // Информация и космос. – 2018. – №1. – С. 81-85.
2. Ullah S., Xuefeng Zh., Feng Zh. TCloud: A Dynamic Framework and Policies for Access Control across Multiple Domains in Cloud Computing // International Journal of Computer Applications (0975–8887). – 2013. – Vol. 62, no.2. – Pp. 1-7.
3. Wu R., Zhang X., Ahn G.-J. et al. ACaaS: Access Control as a Service for IaaS Cloud. – URL: http://sefcom.asu.edu/publications/design-implementation-access-science2013.pdf (Дата обращения: 10.08.2018).

4. Patel S.Ch., Umrao L.S., Singh R.Sh. Policy-based Access Control in Cloud Computing // Proceedings of the International conference on Artificial Intelligent and Soft Computing. – 2012.
5. Meghanathan N. Review of Access Control Models for Cloud Computing // Proceedings of the Third International Conference on Computer Science, Engineering & Applications, ICCSEA, Chennai, India, May 24-26, 2013. – Pp. 77-85.
6. Majumder A., Namasudra S., Nath S. Taxonomy and Classification of Access Control Models for Cloud Environments / Z. Mahmood (ed.) // Continued Rise of the Cloud, Computer Communications and Networks. – 2014. – Pp. 23-32.
7. Frank M., Buhmann J.M., Basin D. On the Definition of Role Mining // Proceedings of the 15th ACM symposium on access control models and technologies (Pittsburgh, PA, USA, June 09 - 11, 2010). SACMAT '10. – ACM, New York, NY, 2010. – Pp. 35-44.
8. Vaidya J., Atluri V., Guo Q. The Role Mining Problem: Finding a Minimal Descriptive Set of Roles // Proceedings of the 12th ACM symposium on Access control models and technologies (Sophia Antipolis, France, June 20 - 22, 2007). SACMAT '10. – ACM, New York, NY, 2007. – Pp.175-184.
9. Saenko I., Kotenko I. Genetic Algorithms for Role Mining Problem // Proceedings of the 19th International Euromicro Conference on Parallel, Distributed and Network-Based Processing (Ayia Napa, Cyprus, February 9-11, 2011). PDP'2011. –IEEE, 2011. – Pp. 646-650.
10. Saenko, I., Kotenko I. Design and Performance Evaluation of Improved Genetic Algorithm for Role Mining Problem // Proceedings of the 20th International Euromicro Conference on Parallel, Distributed and Network-based Processing (Garching, Germany, February 15 - 17, 2012). PDP'2012. – IEEE, 2012. – Pp. 269-274.
11. Jafari M., Chinaei A.H., Barker K., Fathian M. Role Mining in Access History Logs // International Journal of Computer Information Systems and Industrial Management Applications. – 2009. – Vol. 1, no. 1. – Pp. 258-265.
12. Uzun E., Atluri V., Lu H., Vaidya J. An Optimization Model for the Extended Role Mining Problem // Proceedings of the 25th Annual IFIP WG 11.3 Conference (Richmond, VA, USA, July 11 - 13, 2011). DBSec 2011. – Springer, LNCS, vol. 6818, 2011. – Pp.76–89.
13. Blundo C., Cimato S. Constrained Role Mining // Security and Trust Management. Proceedings of the 8th International Workshop (Pisa, Italy, September 13 - 14, 2012), Revised Selected Papers. STM 2012. – Springer, LNCS, vol. 7783, 2012. – Pp.289-304.
14. Such J.M., Criado N. Resolving Multi-party Privacy Conflicts in Social Media // IEEE Transactions on Knowledge and Data Engineering. – 2016. –Vol. 28, no. 7. – Pp. 1851–1863.
15. Such J.M., Rovatsos M. Privacy Policy Negotiation in Social Media // Journal ACM Transactions on Autonomous and Adaptive Systems (TAAS). – 2016. – Vol. 11, no. 1. – Article No. 4.

I.B. Saenko (Saint-Petersburg institute for informatics and automation of Russian Academy of Sciences, Saint-Petersburg), **V.I. Komashinskiy** (The Institute of Transport Problems of Russian Academy of Sciences, Saint-Petersburg)

The system of indices and criteria for evaluation of quality of access control policies in cloudy infrastructures of crucial information systems

The paper considers issues of creating the system of indices and criteria for evaluation of quality of access control policies applied in cloudy infrastructures of crucial information systems. The characteristic of the modern access control models for cloudy infrastructures is provided. Requirements are justified and the formal description for measures and criteria of quality of access policies is done.

ГОРЯЧЕЕ ПОДКЛЮЧЕНИЕ ВНЕШНИХ ЭЛЕКТРОННЫХ КОМПОНЕНТОВ К УСТРОЙСТВАМ ПЛАТФОРМЫ ARDUINO

А. Д. Кулаков[1],
(*Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики (Университет ИТМО),
197101, С.-Петербург, Кронверкский проспект, д.49,
+7(812)232–97–04, kulakov_add@mail.ru*)

Д. С. Левшун[2]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН),
199178, С.-Петербург, 14-я линия В.О., д. 39,
(812) 328-71-81, levshun@comsec.spb.ru*)

Аннотация

В данной работе представлен алгоритм горячего подключения внешних электронных компонентов к устройствам на основе микроконтроллеров. Под горячим подключением понимается изменение конфигурации устройства для взаимодействия с новым набором внешних электронных компонентов. При этом смена конфигурации происходит без загрузки новой прошивки на устройство или его отключения от сети. Наличие подобного функционала позволит на основе специальных запросов динамически изменять функциональные возможности устройств на основе микроконтроллеров. Применение алгоритма продемонстрировано на устройствах платформы Arduino, объединенных в сеть на основе протокола I2C, задействованных в качестве аппаратных интерфейсов комплексной системы киберфизической безопасности.

Ключевые слова: киберфизическая система, горячее подключение, динамическая конфигурация, микроконтроллер.

Введение

Как правило, современные киберфизические системы содержат устройства на основе микроконтроллеров для взаимодействия с физической средой объекта мониторинга [1]. Наличие подобных устройств позволяет системам получать актуальные данные для измерения, сигнализации и управления (например, данные о давлении, температуре, концентрации определенного газа, скорости протекания реакции, значении напряжения или электрического тока в цепи и т.п.). Однако, для обеспечения подобного потока данных на устройства необходимо загружать специальную прошивку для взаимодействия с конкретным набором внешних электронных компонентов. При этом процесс изменения набора компонентов для каждого отдельного устройства связан с его полным отключением от системы, загрузкой новой прошивки и только затем обратным подключением к системе.

Для решения данной проблемы, авторами работы был разработан алгоритм горячего подключения внешних электронных компонентов к устройствам на основе микроконтроллеров. Применимость алгоритма была продемонстрирована на устройствах платформы Arduino [2], объединенных в сеть на основе протокола I2C [3], задействованных в качестве аппаратных интерфейсов комплексной системы киберфизической безопасности [4].

Алгоритм горячего подключение внешних электронных компонентов

В данной работе под горячим подключением понимается изменение конфигурации устройства на основе микроконтроллера для взаимодействия с новым набором внешних электронных компонентов без загрузки новой прошивки и/или его отключения от сети. При этом алгоритм горячего подключения состоит из следующих шагов: (1) инициализация; (2) реализация; (3) рабочий процесс.

Шаг 1. Очевидно, что процесс изменения конфигурации устройства при горячем подключении нового набора внешних электронных компонентов и нормальный процесс его работы в данный момент не совместимы. Это означает, что для перехода устройства к процессу горячего

[1] Магистр.
[2] Аспирант, младший научный сотрудник.

подключения, извне должен поступить специальный сигнал, инициализирующий данный процесс. Данный сигнал позволит как самому устройству, так и связанным с ним, временно изменить свои рабочие процессы и избежать потерь данных или получения некорректных данных.

Шаг 2. После такого как все затронутые процессом горячего подключения устройства изменили свои рабочие процессы, непосредственно реконфигурируемо устройство анализирует предоставленную конфигурацию. Это необходимо для выявления отличий относительно текущей конфигурации и формирования корректного процесса перехода. При этом для обеспечения корректного перехода от одной конфигурации к другой важно учитывать особенности как аппаратной, так и программной реализации задействованных внешних электронных компонентов. Кроме того, необходимо отключить текущий набор внешних электронных компонентов и обеспечить подключение предоставленного набора.

Шаг 3. После завершения процесса горячего подключения, устройство готово к работе с новым набором внешних электронных компонентов. Это означает, что устройство должно сформировать специальный сигнал, означающий, что процессы инициализации и реализации завершены. Данный сигнал позволит как самому устройству, так и связанным с ним, перенастроить свои рабочие процессы и повторно включить устройство с обновленной конфигурацией в сеть.

Реализация горячего подключение внешних электронных компонентов

Передача данных между аппаратными интерфейсами (подчиненными устройствами шины) и концентраторами (главными устройствами шины) комплексной системы киберфизической безопасности осуществляется по протоколу I2C. При таком соединении, один из аппаратных интерфейсов платформы Arduino выступает в качестве главного микроконтроллера последовательной шины данных, а соединенные с ним микроконтроллеры в качестве подчиненных (см. рис. 1). Алгоритм горячего подключения внешних электронных компонентов представлен на рисунке 2.

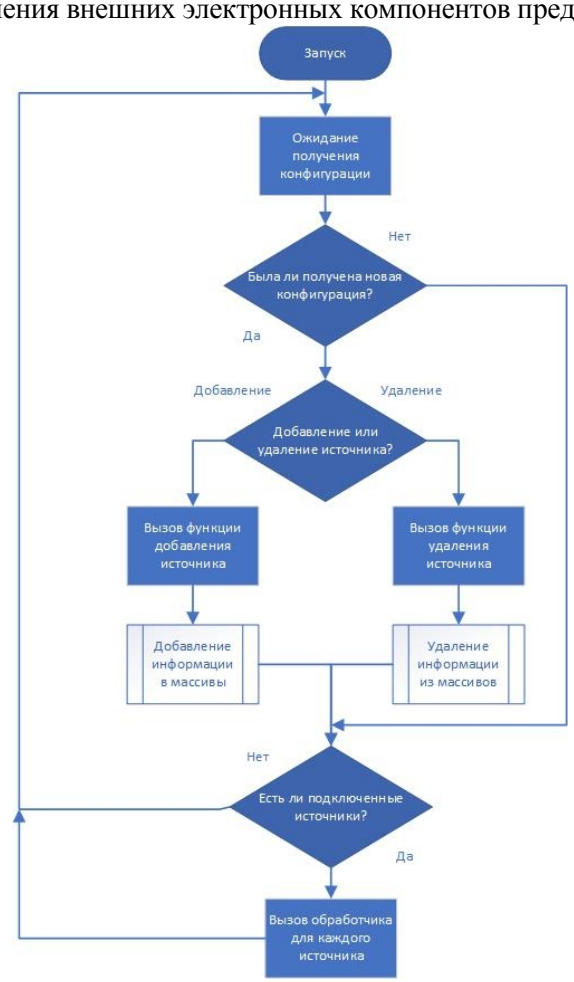


Рис. 1. Алгоритм горячего подключения внешних электронных компонентов

Команда о необходимости изменения конфигурации (как и данные о новой конфигурации) поступают на аппаратный интерфейс с главного устройства сети. Новая конфигурация содержит информацию об отключаемых или подключаемых к устройству внешних электронных компонентах: номера PIN микроконтроллера, необходимых к использованию или освобождению; а также типе подключаемого или отключаемого компонента для проведения процедуры его инициализации или деинициализации соответственно.

После запуска плата ожидает получения новой строки конфигурации. Когда в буфере последовательного порта появляются данные, вызывается специальная функция, которая записывает полученную строку в переменную. В этом случае в бесконечном цикле происходит обработка этой строки и изменение конфигурации платы путем добавления или удаления элементов из массивов. Далее, если после изменения конфигурации все еще есть информация о подключенных источниках, аппаратный интерфейс переходит в состояние готовности к работе с новым набором внешних электронных компонентов. При этом вызываются функции-обработчики для каждого подключенного источника. Если же изменения конфигурации источников не осталось, то плата ожидает дальнейшей смены конфигурации. Нужно отметить, что вышеописанное подключение или отключение внешних электронных компонентов проводилось только на программном уровне, поэтому для полного завершения процесса изменения конфигурации внешние электронные компоненты все ещё необходимо отсоединить от устройства и / или подсоединить к устройству вручную.

Заключение

Принципиальным отличием использования горячего подключения внешних электронных компонентов вместо стандартной эксплуатации заключается в снижении временных затрат на изменение конфигурации устройств системы. При этом процесс изменения конфигурации может быть инициализирован извне (например, на стороне сервера). Отметим, что хотя разработанная программная реализация применима именно для микроконтроллеров платформы Arduino и была написана на языке Wiring [5], данная реализация может быть адаптирована и для других платформ при учете особенностей их аппаратной реализации.

Работа выполнена при поддержке бюджетной темы № АААА-А16-116033110102-5.

ЛИТЕРАТУРА

1. Десницкий В.А., Чечулин А.А., Котенко И.В., Левшун Д.С., Коломеец М.В. Комбинированная методика проектирования защищенных встроенных устройств на примере системы охраны периметра // Труды СПИИРАН. 2016. Вып. 5(48). С.5-31.
2. Официальный форму разработчиков микроконтроллеров платформы Arduino. – URL: https://www.arduino.cc/ (дата обращения: 30.07.2018).
3. Официальная документация по работе с протоколом I2C на устройствах платформы Arduino. – URL: https://www.arduino.cc/en/Reference/Wire (дата обращения: 30.07.2018).
4. Vasily Desnitsky, Dmitry Levshun, Andrey Chechulin and Igor Kotenko. Design Technique for Secure Embedded Devices: Application for Creation of Integrated Cyber-Physical Security System. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), Vol.7, No.2, June, 2016. P.60-80.
5. Официальный сайт фреймворка Wiring для программирования микроконтроллеров с открытым исходным кодом. – URL: http://wiring.org.co/ (дата обращения: 30.07.2018).

A.D. Kulakov (St. Petersburg National Research University of Information Technologies, Mechanics and Optics (ITMO University), St. Petersburg), **D.S. Levshun** (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), St. Petersburg)

Hot-Plugging of external electronic components to Arduino platform devices

Abstract. In this paper, we present a methodology for organizing of the hot-plugging of external electronic components to devices based on microcontrollers. Hot-Plugging refers to changing

of the configuration of the device to make it possible to interact with a new set of external electronic components. The configuration is changed without downloading a new firmware to the device or disconnecting it from the system. The presence of such a functionality will make it possible, on the basis of special request from the system, to dynamically change the functionality of devices based on microcontrollers. Application of the methodology is demonstrated on Arduino platform devices, networked on the basis of the I2C protocol. This devices are used as hardware interfaces of the integrated cyber-physical security system.

ПОДХОД К РАЗРАБОТКЕ МЕТОДИЧЕСКОГО АППАРАТА ДЛЯ ТЕХНОЛОГИЧЕСКОГО УПРАВЛЕНИЯ ВИРТУАЛЬНЫМИ ВЫЧИСЛИТЕЛЬНЫМИ СЕТЯМИ В КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

А. М. Старков[1]
(*Военная академия связи имени им. Маршала Советского Союза С.М. Буденного,
194064, С.-Петербург, Тихорецкий пр., 3,
+7(995)-996-05-01, kadet58v@mail.ru*)

И. Б. Саенко[2]
(*Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский институт информатики и автоматизации» РАН (СПИИРАН),
199178, С.-Петербург, 14-я линия В.О., 39,
+7(812)-328-7181, ibsaen@mail.ru*)

Аннотация

В докладе раскрываются основные этапы разработанного методического аппарата технологического управления виртуальными локальными вычислительными сетями в корпоративных информационных системах и конкретные примеры его применения.

Ключевые слова: корпоративная информационная система, локальная вычислительная сеть, VLAN, имитационное моделирование.

Введение

В целях повышения эффективности информационного взаимодействия сотрудников организации и для снижения производственных затрат руководители предприятия стараются максимально автоматизировать процессы производственного цикла путем построения корпоративных информационных систем (КИС). Основу построения КИС составляют локальные вычислительные сети (ЛВС) организации. Таким образом, эффективность функционирования КИС и организации в целом напрямую зависит от принятых решений по построению и функционированию ЛВС. Одним из перспективных способов повышения эффективности построения и функционирования ЛВС является применение технологии «виртуальной локальной вычислительной сети» (virtual local area network – VLAN). Технология VLAN позволяет решить не только вопросы повышения информационной безопасности в КИС, но и увеличить своевременность предоставления необходимой информации сотрудникам в процессе динамического изменения структуры предприятия. Однако в настоящее время при проектировании схем построения VLAN применяются эмпирические рекомендации, которые при трансформации структуры предприятия не позволяют в полном объеме обеспечить выполнение показателей функционирования ЛВС. Предлагаемый доклад посвящен раскрытию этапов подхода к разработке методического аппарата технологического управления виртуальными локальными вычислительными сетями в корпоративных информационных системах.

Описание подхода

Место VLAN в системе управления КИС. Согласно ГОСТ Р 52653-2006 и ГОСТ РВ 51987-2002 КИС представляет собой разновидность автоматизированной системы управления (АСУ), доступ к которой имеет ограниченный круг лиц, определенный ее владельцем или соглашением участников этой системы.

Анализ различных источников информации по проблемам исследования подходов к синтезу структуры АСУ показал, что данные системы включают следующие уровни: организационного управления, оперативно-технического управления, технологического управления [1-3].

[1] Адъюнкт.
[2] Доктор технических наук, профессор, ведущий научный сотрудник.

На уровне организационного управления (ОУ) обеспечивается реализация целевых задач функционирования системы КИС путём планирования (разработки и коррекции документов по построению) и управления построением системы.

На уровне оперативно-технического управления (ОТУ) осуществляется контроль состояния и изменение структуры вычислительных сетей, управление качеством предоставления услуг связи во взаимодействии с подсистемами организационного и технологического управления.

На уровне технологического управления (ТУ) осуществляется контроль и изменение технического состояния сетевых элементов вычислительной сети (ВС).

Несмотря на определённые каждому уровню целевые задачи, в настоящее время в системах управления все чаще происходит слияние и интеграция функциональных задач различных уровней между собой [1,2,4]. Это, в свою очередь, затрудняет рассмотрение инновационных решений в рамках одного уровня.

Таким образом, для исключения двойственных толкований места технологии VLAN в соответствии с вышеизложенной классификацией уровней АСУ требуется ее более детальное рассмотрение. Под VLAN понимается технология логического сегментирования физически единой локальной вычислительной сети (ЛВС) на множество широковещательных доменов и предоставления пользователям доступа к сетевым ресурсам в соответствии с их принадлежностью к определённым доменам [5,6,7]. Основу данной технологии составляют управляемые коммутаторы, которые обеспечивают разделение сетевого ресурса на основе неполнодоступной схемы связи портов.

В связи с тем, что основой построения КИС является ЛВС, то в зависимости от эффективности принятых решений по построению, функционированию и управлению ЛВС обеспечивается выполнение требований, предъявляемых к КИС предприятия.

Таким образом, в соответствии со своей областью применения технология VLAN относится к уровню ТУ ЛВС предприятия, и выбранная организационная структура построения VLAN оказывает непосредственное влияние на эффективность функционирования ЛВС предприятия.

Требования и критерии, предъявляемые к ТУ ЛВС. Процесс ТУ представляет собой непрерывный циклический процесс по обеспечению целенаправленной деятельности управляемой ЛВС, основные этапы которого отображены на рисунке 1.

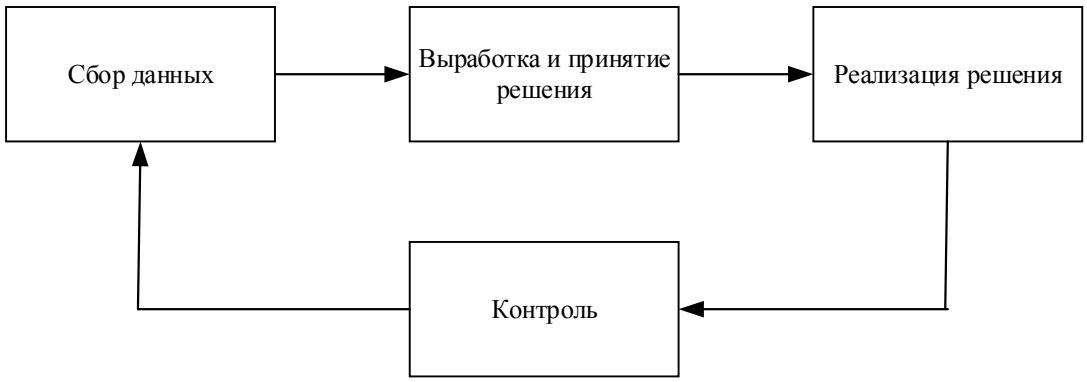


Рис. 1. Взаимосвязь процессов ТУ ЛВС

На этапе сбора данных формируется структура задачи ЛВС на основе изучения руководящих документов, а также проводится синтез описания текущего состояния на основе множества факторов, определяющих построение и функционирование ЛВС.

На этапе выработки решения и принятия решения определяются требования к формируемой (реструктуризируемой) ЛВС, порядок использования, структура ЛВС и ее элементов, а также способы функционирования и восстановления ЛВС.

На этапе реализация решения производится построение (реструктуризация) ЛВС в соответствии с принятым решением и требованиями к ЛВС.

На этапе контроля производится непрерывная проверка информации состояния ЛВС и ее элементов с последующей оценкой рассогласованности текущего и требуемого состояния ЛВС.

ЛВС как объекту управления присущи следующие особенности: сложность, большая размерность и территориальная распределенность, динамичность структуры, ограниченность ресурсов для их построения, а также предъявляемые жесткие требования к процессу управления ЛВС.

Таким образом, при выборе критериев и показателей необходимо исходить из системного подхода к оценке эффективности как ЛВС в целом, так и её составных элементов (функциональных подсистем, процессов управления, устройств). Его результатом должна быть система взаимоувязанных и взаимообусловленных критериев и показателей.

Так как по своему содержанию ТУ представляет собой процесс решения множества управленческих задач, то к нему в полной мере относятся общие требования, предъявляемые к управлению, т.е. устойчивость, непрерывность, оперативность, скрытность и обоснованность принимаемых решений.

В свою очередь ТУ ЛВС представляет собой управление отдельными сетевыми элементами (коммутаторами, маршрутизаторами, мультиплексорами и т.п.), которое влияет на выполнение требований, предъявляемых в целом к телекоммуникационной системе по готовности, устойчивости, мобильности, пропускной способности, разведывательной защищенности, доступности и управляемости.

С другой стороны ЛВС является составной частью АСУ. Соответственно, предъявляемые требования по оперативности, надежности и гибкости справедливы и для ТУ ЛВС.

Проведенный анализ показал, что к ТУ ЛВС ПУ ОО предъявляются требования, изложенные в таблице 1.

Таблица 1

Требования и критерии, предъявляемые к ТУ ЛВС

Свойства	**Показатель**	**Критерий**
Готовность	$P(t_г \le T_г^{доп})$ – вероятность выполнения требуемого комплекса мероприятий за установленное время	$P_г \le P_г^{доп}$
Мобильность	$P(t_{вз} \le T_{вз}^{зад})$ – вероятность своевременного выполнения задачи по изменению структуры	$P_{изм.стр.} \ge P_{изм.стр.}^{доп}$
Пропускная способность	$P(t_\lambda \le T_\lambda^{зад})$ – вероятность обмена заданными потоками сообщений в установленные сроки и с требуемым качеством	$P_\lambda \le P_\lambda^{доп}$, при $Q_\lambda \le Q_\lambda^{треб}$
Управляемость	$P(t_{пер} \le T_{пер}^{зад})$ – вероятность перевода ЛВС из одного состояния в другое(требуемое) за заданное время	$P_{упр} \ge P_{упр}^{треб}$
Доступность	$P(t_{дос} \le T_{дос}^{зад})$ – вероятность доступа пользователей к ресурсу (услугам) ЛВС за заданное время	$P_{дос} \ge P_{дос}^{треб}$
Оперативность	$P(t_{упр} \le T_{упр}^{доп})$ – вероятность того, что продолжительность управления не превысит допустимое время	$P(T_{упр} \le T_{упр}^{доп})$
Своевременность	$P(t_{обм} \le T_{обм}^{доп})$ – вероятность своевременного обмена управляющей информацией (УИ)	$\Delta P_{УИ} \le P_{УИ}^{доп}$
Достоверность	Вероятность ошибочного обмена информацией	$P_{ош.инф.} \le P_{ош.инф.}^{доп}$
Безопасность	Коэффициент требуемой доступности к ресурсам ЛВС	$k_{дост}^{треб} \ge k_{дост}^{доп}$
Надежность	Коэффициент готовности ЛВС на определенном информационном направлении	$k_{гот}^{треб} \ge k_{гот}^{доп}$
Гибкость	$P(t_{прис} \le T_{прис}^{зад})$ – вероятность своевременного приспособления ЛВС к изменению обстановки	$P_{прис} \ge P_{прис}^{доп}$

Из таблицы 1 следует, что для эффективного функционирования ТУ ЛВС необходимо выполнение предъявляемых требований в рамках допустимого времени, определенного руководящими документами предприятия. Таким образом, основным свойством процесса ТУ является оперативность управления, количественным показателем которого выступает длительность цикла ТУ ($T_{ЦТУ}$):

$$T_{ЦТУ} = t_k + t_{сб} + t_{об} + t_{пр} + t_{фки} + t_д ,$$

где t_k – время контроля элемента системы связи; $t_k = t_{об} + t_{фис}$; $t_{об}$ – время обнаружения несоответствия параметров контроля элемента ЛВС (канала связи, коммутационного оборудования) требуемым (допустимым) значениям; $t_{фис}$ – время формирования информации состояния и её передача на пункт управления элементом ЛВС (администратору ЛВС); $t_{сб}$ –время сбора информации о состоянии системы (элемента, сети); $t_{об}$ – время обработки информации о состояния администратором сети; $t_{пр}$ – время принятия решения по сложившейся ситуации; $t_{фки}$ – время формирования командной информации администратором сети; $t_д$ – время доведения командной информации до элементов сети.

Таким образом, от ЛВС требуется обеспечение эффективного (своевременного и качественного) обмена всеми видами информации (сообщениями) между организационными структурами предприятия, своевременным изменением параметров в соответствии с принятым решением, а также передачи сигналов управления в установленные сроки (близкие к режиму реального времени).

Описание задачи (формальная постановка задачи). В настоящее время перед разработчиками КИС возникает важная задача синтеза VLAN с оптимальной структурой [7,9,10].

При этом требуется оптимизировать структуру VLAN по критериям максимальной пропускной способности ЛВС и по доступности к информационным ресурсам данной сети. Задача является нетривиальной, так как для одних и тех же исходных данных существует огромное количество вариантов построения сети. Как правило, при проектировании структуры сети разработчики руководствуются собственным опытом и интуицией, что не всегда приводит к наилучшим результатам.

Поэтому актуальным является создание такой модели функционирования ВЛВС, в которой разработчик лишь указывал бы желаемые параметры сети, а на выходе получал бы оптимальную структуру VLAN.

В качестве исходных данных для построения модели учитываются следующие факторы:

1) структура ЛВС, характеристики входящих в ее состав технических и программных средств;
2) требования по пропускной способности, безопасности (доступности) ЛВС;
3) перечень задач, решаемых с помощью ЛВС КИС;
4) структура информационно-технического взаимодействия должностных лиц в КИС.

Формальная постановка задачи для синтеза оптимальной структуры VLAN будет иметь следующий вид.

Дано:

$G = \{A, B_{mn}, C\}$ - граф структуры ЛВС предприятия;

$U = \{u_1, ..., u_y\}, y = \overline{1, Y}$ – множество вершин графа, соответствующие множеству сетевых устройств, которые являются источниками и (или) приемниками информации при решении ДЛ КИС функциональных задач;

$B = \{b_{mn}\}$ – множество ребер графа между вершинами u_i и u_j, которые соединяют сетевые устройства;

$C = \{c_{ij}\}$ – пропускная способность ребер графа, которые соединяют сетевые устройства;

$\Lambda = \{\lambda_{ij}\}$– множество информационных потоков между i-м и j-м пользователями ЛВС;

$D = \|d_{ij}\|, d_{ij} \in [0,1]$– матрица доступа между абонентами ЛВС предприятия;

$TP_q = \{P_\lambda^{Tp}, Q_\lambda^{Tp}, K_o^{Tp}\}, q = \overline{1, Q}$ – множество требований к ЛВС предприятия, где P_λ^{Tp} – подмножество требований к значениям вероятностей обмена заданными потоками сообщений в установленные сроки и с требуемым качеством Q_λ^{Tp};

K_o^{Tp} – подмножество требований к значениям коэффициента доступности ДЛ к ресурсам ЛВС предприятия.

Требуется:

1. Разработать модель функционирования VLAN на ЛВС предприятия вида $M = \langle X, V, \Lambda, G \rangle$, где $X = \{x_k\}, k = 1...K$ − множество подсетей VLAN, $V = \{v_n\}, n = 1...N$ − множество выходных характеристик ЛВС предприятия.

2. Разработать методику ТУ VLAN с целью повышения эффективности функционирования ЛВС *Ф(t)* на интервале времени T: $\Phi\left(r_{внеш}(t), r_{внут}(t), y(t), T\right) \xrightarrow{u(t)} \max$, где $r_{внеш}(t) \in R_{внеш}$, $r_{внут}(t) \in R_{внут}$, $u(t) \in U$, − внешние, внутренние возмущающие и управляющие воздействия, действующие на сеть в момент времени t, $y(t) \in Y$ – реакция сети по поиску такого варианта организации подсетей VLAN в ЛВС предприятия X при $\begin{cases} V(X) \geq V_{треб} \\ K_{уст} \Rightarrow \max \end{cases}$.

Методический аппарат решения задачи. Для решения поставленной задачи по поиску оптимальной структуры VLAN на первом этапе требуется произвести сбор трафика, циркулирующего в ЛВС предприятия – множества информационных потоков $\Lambda = \{\lambda_{ij}\}$. Данное множество может быть получено как опытным путем, так и при помощи специального программного обеспечения, позволяющего собирать статистику и производить анализ информационного трафика абонентов локальной вычислительной сети. В качестве программного обеспечения, предназначенного для решения данной задачи, предлагается использовать программу сбора сетевого трафика NetFlow Analyzer. Данное множество в дальнейшем позволит проверить достоверность полученных данных в результате имитационного моделирования и адекватность синтезированной модели. Наиболее подходящими временными отрезками для сбора трафика являются день, неделя и месяц.

В ходе объединения полученного множества информационных потоков $\Lambda = \{\lambda_{ij}\}$ и матрицы доступа между абонентами ЛВС предприятия $D = \|d_{ij}\|$ булева матрица разрешенных информационных потоков $A[n,n]$ компьютерной сети для *n* компьютеров. Если $a_{ij} = 1(i, j = 1,...,n)$, то обмен между компьютерами *i* и *j* разрешен. В противном случае этот обмен невозможен.

Для дальнейшего формирования структуры построения VLAN сформируем матрицу структуры построения VLAN согласно следующему алгоритму. Положим, что мы сформировали в компьютерной сети *k* виртуальных подсетей. Каждая из этих подсетей объединяет два и более компьютера. Зададим распределение компьютеров по подсетям с помощью матрицы $S[n,k]$. Если $S_{ij} = 1(i, j = 1,...,k)$, то компьютер *i* принадлежит подсети *j*. В противном случае подсеть *j* не охватывает компьютер *i*.

Матрица *A* играет роль исходных данных задачи. Матрица *X* является результатом решения задачи. Между булевыми матрицами *A* и *S* существует следующая зависимость:

$$A = S \otimes S^T, \tag{1}$$

где S^T – транспонированная матрица *X*, символ $\otimes$ – обозначает булево матричное умножение, которое является формой матричного умножения, основанной на правилах булевой алгебры.

Булево матричное умножение позволяет получать элементы матрицы A согласно следующему выражению: $a_{ij} = \vee_{j=1}^{n} \left(s_{ij} \wedge s_{ji}\right)$.

Для доказательства выражения (1) обозначим множество компьютеров как $P = \{P_n\}$ и множество виртуальных подсетей $V = \{V_k\}$. Между множеством компьютеров и множеством подсетей имеется отображение $S : P \to V$, между множеством подсетей и множеством компьютеров – отображение $S^T : V \to P$. В результате отображение *A* из множества *P* в множество *V* определяется произведением отображений *S* и S^T. Это доказывает справедливость выражения (1). Матрица *A* всегда является симметричной.

Данная задача является разновидностью задач булевой матричной факторизации (БМФ). Из того, что рассматриваемая задача является разновидностью задач БМФ, следует, что она является NP-полной и означает ее неразрешимость формальными методами в приемлемые сроки. Поэтому для решения такого класса задач требуется разработка эмпирических алгоритмов [11,12].

Для этой цели необходимо ввести критерий оптимизации. Чтобы это сделать, следует обратить внимание на тот факт, что в задаче (1) значение k может быть произвольным. Следовательно, критерием оптимизации рассматриваемой задачи является минимальное значение k при условии полного совпадения A и $S \otimes S^T$.

Выбор метода оптимизации структуры VLAN. Существует достаточно большое количество разнообразных классификаций современных информационных технологий, учитывающих парадигмы рассматриваемой предметной области. Применительно к области управления ЛВС можно предложить классификацию, представленную на рисунке 2 [13,14,15].

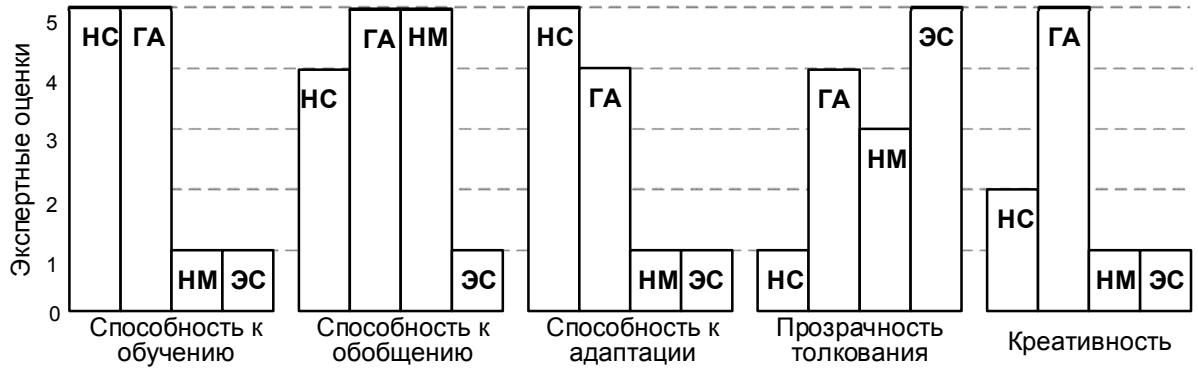


Рис. 2. Сравнение экспертных оценок интеллектуальных систем: НС – нейронные сети; ГА – генетические алгоритмы; НМ – нечеткие системы; ЭС – экспертные системы

Как показывает анализ существующих методов обработки знаний (экспертных систем, генетических алгоритмов, нечеткой логики, нейронных сетей), наиболее подходящим для решения данной оптимизационной задачи являются генетические алгоритмы.

Генетический алгоритм – это способ решения задач оптимизации, использующий для поиска оптимального решения принципы эволюционного развития. Основной механизм эволюции обладает двумя наиболее характерными чертами. Во-первых, новые особи, появляющиеся в некоторой популяции, наследуют свойства своих родителей, причем в равных долях (механизм генетического скрещивания и наследования). Во-вторых, с некоторой вероятностью в течение своей жизни особи могут изменять свои свойства (гены), т.е. мутировать. Эти черты положены в основу генетических алгоритмов оптимизации (ГАО).

ГАО представляет собой метод параллельного поиска глобального экстремума, основанный на использовании в процессе поиска сразу нескольких закодированных соответствующим образом точек (вариантов решения), которые образуют развивающуюся по определенным случайным законам популяцию [14,16,17].

Главное отличие ГАО от традиционных методов поиска оптимального решения состоит в том, что на каждом своем шаге вычислений, данный алгоритм имеет не с одним, а с несколькими значениями вектора оптимизируемых параметров, которые образуют популяцию хромосом.

В основу построения метода решения задачи первоначальной конфигурации VLAN положим генетический алгоритм. Существуют различные взгляды на последовательность и содержание шагов генетического алгоритма [15,16,17].

Предлагается придерживаться следующих шагов.

1. Определение функции пригодности, которая показывает, почему одно из возможных решений задачи мы считаем лучше, чем другое решение.

2. Кодирование возможных решений задачи. Закодированное решение в терминологии генетического алгоритма называется особью. Обычно решения кодируются с помощью символьных или числовых строк. Отдельный символ этого кода называется геном. Совокупность генов в строке называется хромосомой.

3. Формирование исходного множества особей, или популяции. Как правило, этот процесс проходит случайным образом, однако количество особей в популяции N является постоянным. Оценка всех особей в популяции с помощью функции пригодности и сортировка их по убыванию ее значения.

4. Выбор пар особей, которые называются родителями, для формирования новых особей, называемых потомками, путем скрещивания. Хромосомы родителей разбиваются на фрагменты. Затем происходит обмен фрагментами родительских хромосом, чтобы сформировать хромосомы потомков. Новые особи оцениваются с помощью функции пригодности и добавляются в текущую популяцию.

5. Выбор особей для мутации их хромосом. Мутация и оценка этих особей.

6. Селекция популяции, которая заключается в оставлении в ней N особей, обладающих самыми высокими значениями функции пригодности. Остальные особи (самые «плохие») удаляются из популяции.

Если выполняются критерии завершения алгоритма, то в качестве решения оптимизационной задачи выбирается особь с максимальным значением функции пригодности. Иначе происходит возврат к шагу 3.

Для того чтобы использовать генетический алгоритм для решения данной задачи, необходимо определить функцию пригодности и порядок кодирования решений.

В известных работах по применению генетических алгоритмов для решения задачи БМФ функция пригодности строилась только на основании эвклидова расстояния между заданной булевой матрицей и булевого произведения матриц. При полном совпадении этих матриц функция пригодности принимала максимальное значение. Однако в нашем случае евклидово расстояние между заданной булевой матрицей A и булевым произведением матриц S и S^T ($S \otimes S^T$) не является достаточным. По одному этому критерию могут быть получены тривиальные решения, а они являются плохими. Необходимо дополнительно учитывать требование того, чтобы в матрице S количество столбцов k было минимальным.

Программная среда имитационного моделирования. Для апробации полученных результатов, а также проверки различных вариантов построения структуры VLAN целесообразно использовать программный продукт имитационного моделирования [18,19,20].

Наиболее подходящим решением является программная система Riverbed Modeler, которая предоставляет широкие возможности моделирования вычислительной сети, представленной в графическом виде, что является одним из основных преимуществ, так как пользователь имеет возможность видеть, как всю сеть в целом, так и при необходимости отдельные ее участки.

Также можно учитывать размеры модели сети, создавая проекты на шаблоне глобального масштаба, районного, кампусного или размером офиса, учитывая расстояния между узлами. При необходимости имеется возможность использовать схему здания, где размещается или планируется разместить вычислительную сеть, получая готовый план-проект, который можно изменять только в рамках ограниченных средой размещения сети [21].

Еще одним преимуществом данного продукта является, что Riverbed Modeler Academic Edition является абсолютно бесплатной утилитой, предназначенной в образовательных целях для студентов учебных заведений.

Также данный программный продукт позволяет строить модель функционирования ЛВС без привязки к конкретному производителю коммутационного оборудования, и в нем реализована функция VLAN.

Заключение

В докладе раскрыты основные этапы предлагаемого подхода к разработке методического аппарата технологического управления виртуальными локальными вычислительными сетями в корпоративных информационных системах.

Рассмотрены особенности функционирования КИС, выбора наиболее значимых свойств эффективного функционирования ЛВС предприятия и поэтапного функционирования системы управления на технологическом уровне управления.

Сформулирована постановка задачи исследования и выработаны предложения по ее решению, касающиеся алгоритма решения оптимизационной задачи и апробации полученных решений в программной системе Riverbed Modeler.

Работа проводилась при частичной финансовой поддержке РФФИ (проекты 16-29-09482, 18-07-01369 и 18-07-01488) и бюджетной темы № АААА-А16-116033110102-5.

ЛИТЕРАТУРА

1. Арсланов Х.А. Автоматизированная система управления связью в Вооруженных Силах Российской Федерации и приоритетные направления ее развития. [Электронный ресурс] / Х.А. Арсланов, А.С. Башкирцев, А.М. Лихачев // – Режим доступа: http://www.army.informost.ru/2016/pdf/1-5.pdf
2. Арсланов Х. А. Концептуальные основы развития объединенной автоматизированной цифровой системы связи Вооруженных Сил Российской Федерации / Х.А. Арсланов, А.В. Абрамович, А.М. Лихачев // Сборник «Связь в вооруженных силах Российской Федерации – 2014». – С. 18–24.
3. Арсланов Х. А. Актуальные научно-практические проблемы развития ОАЦСС ВС РФ / Х.А. Арсланов, А.М. Лихачев // Сборник «Связь в вооруженных силах Российской Федерации – 2015». – С. 29–36.
4. Филимонов А.Ю. Построение мультисервисных сетей Ethernet. – СПб: БХВ-Петербург, 2007. – 592с.
5. Ермаков А.Е. Основы конфигурирования корпоративных сетей Cisco. М.: ФГБОУ «УМЦ ЖДТ», 2013. — 247 с.
6. Олифер В.Г., Олифер Н.А. Компьютерные сети. Принципы, технологии, протоколы. Учебник для вузов. – СПб.: Питер, 2007. – 960с.
7. Демин И.В. Методы анализа корпоративной вычислительной сети, построенной по технологии VLAN (диссертационное исследование). – М.: Московский государственный институт электроники и математики (технический университет), 2002. – 119 с.
8. Котенко И.В. Теория и практика построения автоматизированных систем информационной и вычислительной поддержки процессов планирования связи на основе новых информационных технологий. – СПб.: ВАС, 1998. – 403 с.
9. Саенко И.Б., Котенко И.В. Применение средств генетической оптимизации и визуального анализа для формирования схем доступа в ВЛВС // Информационные технологии и вычислительные системы. – 2015. – №1. – С. 33-46.
10. Saenko I., Kotenko I. Design of Virtual Local Area Network Scheme based on Genetic Optimization and Visual Analysis // Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA). – Vol.5, no.4. –2014. – Pp.86-102.
11. Miettinen P. Dynamic Boolean Matrix Factorizations // Proceedings of the 2012 IEEE 12th International Conference on Data Mining, ACM, New York, 2012. – Pp. 519–528.
12. Батищев Д.И. Генетические алгоритмы решения экстремальных задач. – Учебное пособие. – Нижний Новгород, 1995. – 63 с.
13. Терехов В.И. Применение гибридных систем вычислительного интеллекта для выбора рационального варианта управленческого решения // Военная мысль. – 2009. – №3. – С. 30–34.
14. Макаров И.М. Искусственный интеллект и интеллектуальные системы управления / И.М. Макаров, В.М. Лохин, С.В. Манько, М.П. Романов; [отв. ред. И.М. Макаров]; Отделение информ. технологий и вычислит. систем РАН. – М.: Наука, 2006. – 333 с.
15. Фомин Л. А., Будко П. А. Эффективность и качество инфокоммуникационных систем. Методы оптимизации: Монография. – Москва: ООО ИФ «Физико-математическая литература», 2008. – 296 с.
16. De Jong K.A. Analysis of the behavior of a class of genetic adaptive systems, Ph. D. Thesis – University of Michigan, Ann Arbor, 1975.
17. Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы./ Под ред. В.М. Курейчика – 2-е изд., испр. и доп. – М. ФИЗМАЛИТ, 2006.
18. Клейнрок Л. Теория массового обслуживания. – М.: Машиностроение, 1979. – 432 с.
19. Клейнрок Л. Вычислительные системы с очередями – М.: Мир, 1979. – 600 с.
20. Бушуев С.Н., Осадчий А.С., Фролов В.М. Теоретические основы создания информационно-технических систем. – СПб: ВАС, 1998. – 404 с.
21. Тарасов В.Н. Проектирование и моделирование сетей связи в системе Riverbed Modeler. Лабораторный практикум. / В.Н. Тарасов, Н.Ф. Бахарева, С.В. Малахов, Ю.А. Ушаков. – Самара: 2016. – 260 с.

A.M. Starkov (The Military academy of telecommunications named after Marshal of the Soviet Union S.M. Bydyonny, Saint-Petersburg), **I.B. Saenko** (Saint-Petersburg institute for informatics and automation of Russian Academy of Sciences, Saint-Petersburg)

An approach to developing methodological foundations of technological management of virtual computer networks in corporate information systems

The paper describes the main stages of the developed methodological foundations of technological management of virtual computer networks in corporate information systems and specific examples of its application. The problem statement of a research is formulated. Proposals on solving the problem and including an algorithm of solving the optimization task and approbation of the received solutions are discussed.

РЕАЛИЗАЦИЯ ПОЛУНАТУРНОЙ МОДЕЛИ ЖЕЛЕЗНОЙ ДОРОГИ ДЛЯ АНАЛИЗА ИНЦИДЕНТОВ БЕЗОПАСНОСТИ

Ю. Е. Бахтин[1], **Д. С. Левшун**[2]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), 199178, С.-Петербург, 14-я линия В.О., 39, (812) 328-71-81, yurii.bakhtin@gmail.com, levshun@comsec.spb.ru*)

Аннотация

Широкое распространение получили комплексные системы мониторинга и управления передвижением железнодорожного транспорта. Для проведения оценки методов защиты от киберфизических атак авторами разрабатываются аналитические и полунатурные модели этих систем. В качестве наработок для построения модели используется стенд железной дороги в масштабе 1:87, на основе которого реализуется система мониторинга и управления передвижением железнодорожного транспорта с целью подключения к аналитической модели и проведения анализа инцидентов безопасности.

Ключевые слова: железная дорога, модель, DCC, киберфизическая система.

Введение

Широкое распространение получили комплексные системы мониторинга и управления передвижением железнодорожного транспорта, такие как Positive Train Control [1] и European Rail Traffic Management System (ERTMS) [2], в частности European Train Control System (ETCS) или Комплексное локомотивное устройство безопасности (КЛУБ-У) ижевского радиозавода [3]. Системы управления инфраструктурой железнодорожного транспорта позволяют в реальном времени выполнять комплекс задач управления, в том числе проводить мониторинг состояния отдельных компонентов системы, отслеживать местоположение локомотивов, управлять стрелками, светофорами, автоматическими шлагбаумами и т.п.

Отказы подобных систем могут привести к большим экономическим потерям и даже катастрофам. Поэтому подобные системы обязательно оснащаются аварийно-предупредительной сигнализацией, которая предназначена для оповещения оператора о неполадках в системе. Однако, с повышением автоматизации управления железнодорожным сообщением, все большую угрозу представляют именно киберфизические атаки злоумышленников, и, соответственно, все более актуальными становятся способы противодействия им [4].

В таких критических системах, как инфраструктура железнодорожного транспорта, важно как можно быстрее обнаружить и устранить имеющиеся угрозы безопасности [5]. При этом проведение экспериментов на реальной инфраструктуре железнодорожного транспорта не представляется возможным, поэтому широкое распространение получили аналитические и полунатурные модели.

Системы мониторинга и управления передвижением железнодорожного транспорта

В общем виде, система мониторинга и управления передвижением железнодорожного транспорта представляет собой комплексную систему, структурный состав которой можно представить в виде четырех основных сегментов: (1) сегмент контроля и анализа; (2) среда передачи данных; (3) сегмент локомотивов и (4) придорожный сегмент.

В сегменте контроля и анализа осуществляется отображение состояния объектов и результатов диагностирования системы, текущее местоположение и скорость поездов, а также хранение параметров управления (временных ограничений скорости и другой информации). Основной задачей операционного центра является анализ текущего состояния железнодорожной инфраструктуры, управление временным графиком поездов и дистанционное управление скоростью локомотивов.

[1] Младший научный сотрудник.
[2] Аспирант, младший научный сотрудник.

Среда передачи данных представляет собой беспроводную телекоммуникационную сеть. Данный сегмент предназначен для обеспечения оперативной и надежной передачи данных между всеми компонентами комплексной системы.

Сегмент локомотивов включает в себя бортовые компьютеры поездов, тормозные механизмы, а также органы управления движением. В данном сегменте осуществляется непосредственное управление движением локомотива. Кроме того, данный сегмент отвечает за прием информации от спутника о текущем местоположении локомотива, оперативной отправке информации о текущей скорости и местоположении в операционный центр и прием управляющих команд.

Придорожный сегмент представляет собой совокупность исполнительного и контрольного оборудования на железнодорожном пути. Данное оборудование включает в себя приводы и интерфейсные блоки стрелок, светофоры, устройства рельсовых цепей, автоматические шлагбаумы, а также различные датчики и сенсоры.

Архитектура полунатурной модели железной дороги

Для проведения экспериментов по применению методов защиты необходимо разработать аналитическую и полунатурную модели системы управления железнодорожным сообщением. При этом полунатурная модель может быть разработана на основе готовых решений – моделей железных дорог под цифровым управлением. Цифровое управление локомотивами позволит программным путем имитировать работу машиниста как при взаимодействии с оператором/диспетчером, так и в случае, когда никаких команд от оператора не поступает или связь с ним нарушена.

Архитектура разработанной полунатурной модели системы управления железнодорожным сообщением состоит из следующих блоков: (1) блок сбора данных; (2) блок передачи команд, (3) блок аналитики; (4) блок интерпретации команд; (5) блок графического пользовательского интерфейса; (6) блок системы управления; (7) база данных (см. рис. 1).

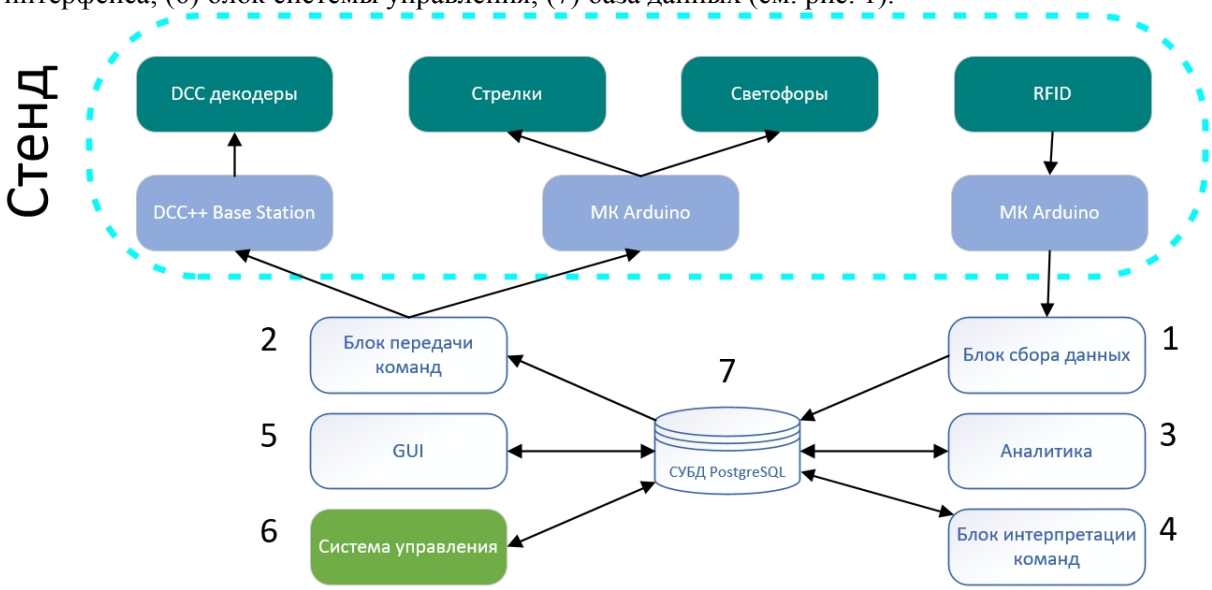


Рис. 1. Архитектура полунатурной модели

Блок сбора данных собирает с макета данные о текущем местоположении и скорости передвижения локомотивов, и, совместно с блоком аналитики, обеспечивает обратную связь для системы управления.

Блок передачи команд обнаруживает поступление новых команд из системы управления и графического пользовательского интерфейса и передает их соответствующему адресату – локомотиву или контроллеру инфраструктуры.

Блок аналитики решает задачу по обработке поступающей информации: расчёту скорости движения составов, и внесению в базу данных полученных результатов, на который опирается система управления при контроле работы модели.

Блок интерпретации команд отвечает за обработку поступающих из внешней системы управления команд, после чего их сможет использовать блок передачи команд. Необходимость интерпретации команд связана с тем, что управление моделями поездов производится командами, связанными с тяговым/тормозным усилием, а поэтому требуется приведение реальных команд к протоколу DCC с учётом характеристик отдельных локомотивов.

Блок графического пользовательского интерфейса обеспечивает удобство взаимодействия с объектами на стенде: позволяет осуществлять ручное управление моделью при отсутствии системы управления или её некорректной работе. Кроме того, этот блок позволяет смоделировать работу машиниста/диспетчера или инцидент, который должна разрешить система управления

Блок системы управления представляет собой внешний блок и отвечает за управление расписанием движения железнодорожного транспорта. Интеграция с данным блоком происходит через базу данных, на основе содержимого которой блок получает информацию о состоянии системы и возвращает управляющие команды.

База данных объединяет различные блоки системы, а также предоставляет возможность разнести функциональность на разные машины, тем самым упрощая внесение изменений в модель – при необходимости можно быстро изменить схему взаимодействия блоков.

Реализация полунатурной модели железной дороги

Полунатурная модель выполнена с архитектурой аналогичной Positive Train Control и European Railway Traffic Management System, с явной средой передачи данных (в модели ей выступает база данных), подсистемой определения местоположения, подсистемой контроля стрелок и прочей инфраструктуры, и непосредственно системой управления.

Реализация полунатурной модели построена с использованием моделей железных дорог под цифровым управлением компании PIKO в масштабе H0 (1:87). Управление моделями железных дорог осуществляется по разработанному национальной ассоциацией железнодорожных моделей (NMRA) протоколу Digital Command Control (DCC) [6]. База данных реализована на основе СУБД PostgreSQL [7]. В модели используется макет с размерами 4x1.5м с несколькими кольцевыми путями, представленный на рисунке 2.

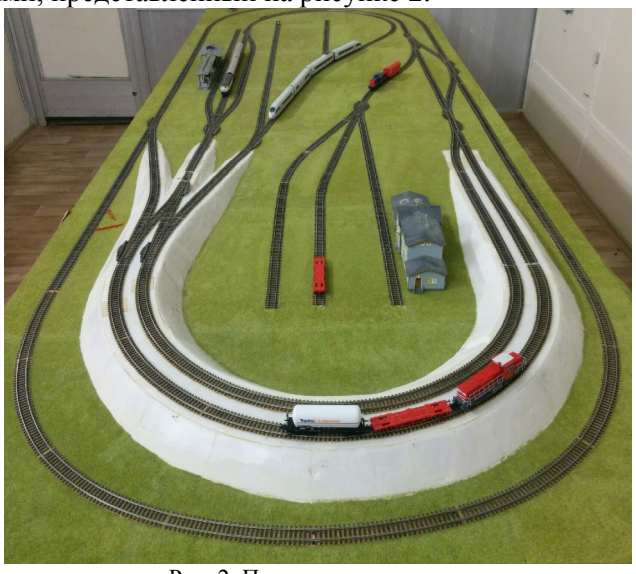


Рис. 2. Полунатурная модель

DCC декодеры в данной модели только мобильные — локомотивы, но потенциально в схеме могут присутствовать и стационарные декодеры, например, для управления стрелками. Из-за недостаточной мощности для управления несколькими поездами и невозможности автоматизированного управления была произведена замена базовой станции PIKO на DCC++ BaseStation [8] на основе популярного микроконтроллера Arduino Mega [9] и H-моста BTS 7960 для выдачи управляющих сигналов высокого напряжения и мощности.

Сложность интеграции со сторонним программным обеспечением и высокая стоимость вынуждают отказаться от готовых контроллеров для периферии, а собственноручная реализация стационарных DCC декодеров накладывает лишние ограничения этого протокола, поэтому было принято решение разработать собственную систему управления стрелками, светофорами и прочей периферией.

На первых этапах управление моделью осуществлялось в программном обеспечении Java Model Railroad Interface (JMRI) [10]. Это гибко настраиваемый комплекс инструментов управления железнодорожными моделями, но по мере уточнения требований к разрабатываемой модели гибкость настройки стороннего программного обеспечения перестала быть достаточной. Основная проблема заключалась в ограничении возможности по организации обратной связи с локомотивами при идентификации и определении местоположения состава. Обычно для определения занятости участка железной дороги используются магнитные датчики, но они не позволяют отличать локомотивы друг от друга. Удовлетворяющая требованиям технология – это радиочастотные метки (RFID), но подключить считыватели меток к JMRI и на их основе вычислять местоположение и скорость поезда оказалось невозможным.

Реализация блока сбора данных опирается на работу RFID считывателей на критичных участках дороги, радиометок на локомотивах и передачу информации на центральный сервер. Сначала в модели использовались RFID считыватели с частотой 125 кГц, однако скорость их работы была недостаточной для обработки меток на высоких скоростях. В связи с чем были выбраны считыватели MFRC522, работающие на частоте 13.56 МГц. В связке считыватель-микроконтроллер метка распознавалась корректно, однако дальнейшие эксперименты показали, что работа нескольких считывателей на одной шине приводит к периодическим сбоям. Это связано с тем, что полное считывание идентификатора метки занимает порядка 43 мс, в течение которых другие считыватели не могут передавать информацию в шину данных. При этом считыватели не имеют внутренней памяти, а потому не способны хранить результаты считывания, что в итоге приводило к потере данных. Решением данной проблемы стало использование микроконтроллеров Arduino Pro Mini для работы с RFID-считывателями, а также последующее их объединение в общую шину на основе протокола I2C [11] (см. рис 3).

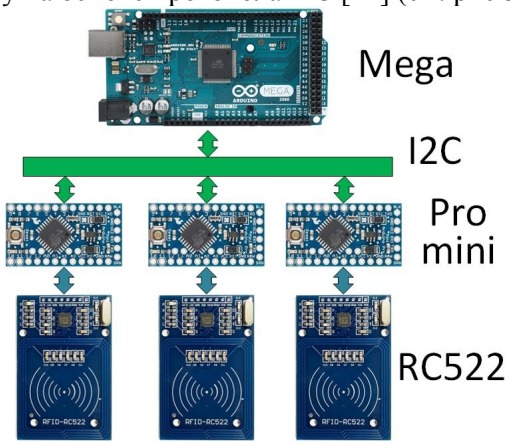


Рис. 3. Схема подключение RFID считывателей

Заключение

В результате работы была реализована полунатурная модель железной дороги с обратной связью. Дальнейшие исследования направлены на создание методики по интерпретации управляющих команд реальных локомотивов, на прогнозирование местоположения локомотивов, исходя из заданной им скорости и посчитанной с помощью радиометок, с целью как можно более полного согласования с аналитической моделью управления расписанием движения поездов. Комплексной целью будущих исследований является моделирование инцидентов безопасности на основе подходов, предложенных в [12-14].

Работа выполнена при поддержке бюджетной темы № АААА-А16-116033110102-5.

ЛИТЕРАТУРА

1. Официальный сайт ассоциации американских железных дорог. – URL: https://www.aar.org/article/freight-rail-positive-train-control/ (дата обращения: 30.07.2018).
2. Официальный сайт европейской системы управления железнодорожным трафиком. – URL: http://www.ertms.net/ (дата обращения: 30.07.2018).
3. Официальный сайт ижевского радиозавода, комплексное локомотивное устройство безопасности. – URL: https://www.irz.ru/products/20/index.htm (дата обращения: 30.07.2018).
4. Котенко И.В., Чечулин А.А., Левшун Д.С. Анализ защищенности инфраструктуры железнодорожного транспорта на основе аналитического моделирования // Защита информации. Инсайд, № 6(78), 2017. С.48-57.
5. Федорченко А.В., Левшун Д.С., Чечулин А.А., Котенко И.В. Анализ методов корреляции событий безопасности в SIEM-системах. Часть 2 // Труды СПИИРАН. 2016. Вып. 6(49). С.208-225.
6. Официальный сайт национальной ассоциации железнодорожных моделей, протокол DCC. – URL: https://www.nmra.org/dcc-rps-standards (дата обращения: 30.07.2018).
7. Свободно распространяемая объектно-реляционная система управления базами данных PostgreSQL. – URL: https://www.postgresql.org/ (дата обращения: 30.07.2018).
8. Официальный сайт проекта базовой станции для управления локомотивами по протоколу DCC++. URL: https://sites.google.com/site/dccppsite (дата обращения: 30.07.2018).
9. Официальный форму разработчиков микроконтроллеров платформы Arduino. – URL: https://www.arduino.cc/ (дата обращения: 30.07.2018).
10. Официальный сайт проекта по разработке инструментов компьютерного управления моделями железных дорог Java Model Railroad Interface. – URL: http://www.jmri.org/ (дата обращения: 30.07.2018).
11. Официальная документация по работе с протоколом I2C на устройствах платформы Arduino. – URL: https://www.arduino.cc/en/Reference/Wire (дата обращения: 30.07.2018).
12. Kotenko I., Chechulin A.. Computer Attack Modeling and Security Evaluation based on Attack Graphs // The IEEE 7th International Conference on “Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications” (IDAACS'2013). Proceedings. Berlin, Germany, September 12-14, 2013. P.614-619.
13. Kotenko I., Konovalov A., Shorov A. Agent-based simulation of cooperative defence against botnets // Concurrency Computation Practice and Experience. 2012. Vol.24. № 6. P.573-588.
14. Kotenko I., Stepashkin M. Network Security Evaluation based on Simulation of Malefactor's Behavior // SECRYPT 2006 - International Conference on Security and Cryptography, Proceedings International Conference on Security and Cryptography, SECRYPT 2006. sponsors: IBM, Polytechnic Institute of Setubal. Setubal, 2006. P.339-344.

Y.E. Bakhtin, D.S. Levshun (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), St. Petersburg)

Development of semi-natural railway model for security incidents analysis

Abstract. Complex systems for monitoring and controlling of railway infrastructure are widely spread. Experiments on the use of cyber-physical security systems for its protection require the development of analytical and semi-natural models of such systems. To build the semi-natural model, the railway stand is used in a scale of 1:87. On its basis, the system for monitoring and controlling the movement of railway transport is implemented to connect the semi-natural model to the analytical model for further analysis of security incidents.

ОПЕРАТИВНОЕ УПРАВЛЕНИЕ МАРШРУТИЗАЦИЕЙ В МУЛЬТИСЕРВИСНЫХ СЕТЯХ СВЯЗИ С УЧЕТОМ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ И СЕТЕВОЙ БЕЗОПАСНОСТИ ПРИ НЕПОЛНОТЕ КОНТРОЛИРУЕМОЙ ИНФОРМАЦИИ

С. А. Агеев[1], **И. Б. Саенко**[2]
(*Санкт-Петербургский институт информатики и автоматизации
Российской академии наук (СПИИРАН),
199178, С.-Петербург, 14-я линия В.О., 39,
(812) 328-71-81, serg123–61@mail.ru, ibsaen@mail.ru*)

Аннотация

В докладе предлагается новый подход к решению проблемы адаптивной маршрутизации в мультисервисных сетях связи, основанный на совместном применении метода многопутевой маршрутизации потоков данных и интегрального критерия качества, который формируется на основании анализа оценки нечеткой иерархической ситуации текущего состояния сети. Оценка формируется по результатам постоянно проводимого мониторинга сети с помощью методов нечеткой классификации и нечеткого логического вывода. В отличие от известных методов и протоколов адаптивной маршрутизации в предлагаемом подходе дополнительно учитывается уровень информационной безопасности, а также техническое состояние сетевых элементов и вероятность потери пакетов.

Ключевые слова: мультисервисная сеть связи, многопутевая маршрутизация, нечеткий логический вывод, маршрутная метрика, индикаторная функция.

Введение

Создание современных мультисервисных сетей связи (МСС) представляет собой сложную многокритериальную задачу. Сложность этой задачи обусловлена следующими факторами: большой распределенностью и неоднородностью МСС; их функционированием в условиях априорной неопределенности относительно внешней среды; многообразием и высокой динамикой физических и информационных процессов, протекающих в МСС.

Особое место при проектировании и при эксплуатации МСС занимает задача обеспечения информационной безопасности в сети. Существенной особенностью решения этой задачи в современных условиях является необходимость учета угроз реализации и воздействия на МСС достаточно большого количества различных типов компьютерных атак. Успешное решение возможно при условии разработки и практического применения моделей и методов оперативного реагирования МСС на возникающие признаки деструктивных информационных воздействий [1, 2]. В настоящей статье в интересах решения задачи обеспечения информационной безопасности МСС предлагается усовершенствованный метод многопутевой маршрутизации потоков данных. Предлагаемый метод отличается от известных методов своим подходом к формированию интегрального критерия качества выбираемого маршрута.

Следует отметить, что при выборе оптимального маршрута каждый маршрутизатор должен иметь максимально полную информацию о техническом состоянии сети в целом, т.е. о состоянии всех остальных узлов коммутации и состоянии линий связи. Однако в сетях, имеющих большую размерность и высокую динамику изменения технического состояния, эти данные, как правило, поступают в узлы коммутации с запозданием. В результате к моменту принятия решения о корректировке параметров алгоритмов маршрутизации эти данные становятся неактуальными. Следовательно, алгоритмы маршрутизации функционируют в условиях неопределенности текущего и будущего сетевого состояния [3-8]. Это фактор определяет выбор способов и методов управления сетевой маршрутизацией. Предлагаемые в настоящей работе методы, основанные на использовании методов оценки нечеткой сетевой ситуации, нечеткого логического вывода, нечеткой кластеризации и нечеткого математического программирования, позволяют оперативно решить возникающую проблему.

[1] Кандидат технических наук, доцент, ведущий научный сотрудник.
[2] Доктор технических наук, профессор, ведущий научный сотрудник.

Описание подхода к оперативному управлению маршрутизацией в МСС

Постановка задачи исследования. Пусть задана структура МСС, состоящей из N узлов коммутации и M линий связи. Известно решение задачи многопутевой маршрутизации при следующих допущениях [9-15]:

1) все линии связи обладают конечной надежностью, т.е. возможен отказ линии связи;

2) все линии связи помехоустойчивы;

3) время обработки в узлах коммутации пренебрежимо мало по сравнению со временем других процессов, протекающих в сети;

4) узлы коммутации имеют бесконечную память (длина очереди на обслуживание заявок бесконечна);

5) длины всех передаваемых сообщений независимы и распределены по показательному закону со средним значением $1/\mu$;

6) трафик, поступающий в сеть, состоит из сообщений, имеющих одинаковый приоритет, и образует пуассоновский поток со средним значением γ_{ij} для сообщений, возникающих в узле i и предназначенных узлу j. Тогда полный внутренний сетевой трафик равен:

$$\gamma = \sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij} \ . \tag{1}$$

Каждая линия связи состоит из единственного дуплексного канала связи с пропускной способностью $d_{k,l}$, где (k, l) – линия связи между узлами k и l, соответственно. Если линия связи между узлами k и l отсутствует, то $d_{k,l} = 0$.

Пусть $z_{k,l}^{(i,j)}$ – доля потока γ_{ij}, проходящая по линии (k, l). Эта величина лежит в диапазоне от 0 до 1. В этом случае будет справедливо следующее соотношение:

$$\lambda_{k,l} = \sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij} z_{k,l}^{(i,j)} . \tag{2}$$

где $\lambda_{k,l}$ – величина потока в линии (k, l), определяемая потоком γ_{ij}.

Для переменных $z_{k,l}^{(i,j)}$ должно выполняться условие сохранения потока в сети, которое, в простейшем случае, имеет следующий вид:

$$\sum_{k=1}^{N} z_{k,l}^{(i,j)} - \sum_{k=1}^{N} z_{l,k}^{(i,j)} = \begin{cases} -1, & \text{если } l = j, \\ 0, & \text{если } l \neq i, j, \\ 1, & \text{если } l = i. \end{cases} \tag{3}$$

Это же условие для локального сетевого узла при условии, что каждая линия связи может быть разделена на m каналов связи, имеет следующий вид:

$$\sum_{i \in m}\sum_{k=1}^{M} z_{k,l}^{(i,j)} - \sum_{j \in m}\sum_{k=1}^{M} z_{l,k}^{(i,j)} = \begin{cases} -1, & \text{если } l = j, \\ 0, & \text{если } l \neq i, j, \\ 1 & \text{если } l = i. \end{cases} \tag{4}$$

где i – узел-источник, j – узел-получатель сообщения; равенства $l = i$ и $l = j$ характеризуют, соответственно, узел-источник и узел-получатель.

Пусть t_{ij} – среднее время, затрачиваемое на передачу сообщения, которое возникло в узле i и предназначается узлу j. Тогда средняя задержка сообщения в сети определяется как взвешенная сумма межконцевых задержек:

$$T_{cp} = \frac{1}{\gamma}\sum_{k=1}^{N}\sum_{l=1}^{N}\lambda_{k\,l}\, t_{k\,l}, \tag{5}$$

где t_{kl} – среднее время пребывания сообщения в линии связи (k, l).

Среднее время пребывания сообщений в линии связи (k, l) равно сумме времени передачи сообщения и времени ожидания в очереди на передачу. Следовательно, справедливо следующее выражение:

$$t_{k\,l} = 1/(\mu\, d_{k\,l} - \lambda_{k\,l}) \ . \tag{6}$$

Введем обозначение: $f_{k,l} = \lambda_{k,l} / \mu_{k,l}$ – величина потока в линии (k, l), выраженная, например, в Кбайт/сек. Тогда справедливо следующее выражение:

$$T_{cp} = \frac{1}{\gamma} \sum_{k=1}^{N} \sum_{l=1}^{N} f_{k,l} / (d_{k,l} - f_{k,l}) \ . \tag{7}$$

Сделанные предположения и введенные обозначения позволяют сформулировать оптимизационную задачу поиска таких значений $z_{k,l}^{(i,j)}$, которые бы обеспечили наименьшее (оптимальное) значение величины T_{cp} [9-11].

ИЗВЕСТНО:

1) топологическая структура МСС;

2) матрица входных потоков $\left\| \gamma_{i,j} \right\|$;

3) пропускные способности линий связи $\left\| d_{k,l} \right\|$;

4) средняя длина сообщений $1/\mu$.

ТРЕБУЕТСЯ ОПРЕДЕЛИТЬ:

1. Потоки в линиях связи $f_{k,l}$ такие, что:

$$T_{cp} = \frac{1}{\gamma} \sum_{k=1}^{N} \sum_{l=1}^{N} f_{k,l} / (d_{k,l} - f_{k,l}) \ \to min \,, \tag{8}$$

при выполнении ограничений:

$$f_{k,\,l} = \frac{1}{\gamma} \sum_{k=1}^{N} \sum_{l=1}^{N} \gamma_{i,\,j}\, z_{k,\,l}^{(i,\,j)} \ , \quad k, \quad l = \overline{1,\,N} \ , \ f_{k,\,l} < d_{k,\,l} \ ; \ k, l = \overline{1,\,N} \,,$$

$$\sum_{k=1}^{N} z_{k,\,l}^{(i,\,j)} - \sum_{k=1}^{N} z_{l,\,k}^{(i,\,j)} = \begin{cases} -1, & \text{если} \quad l = j, \\ 0, & \text{если} \quad l \neq i,\ j, \ 0 \le z_{k,\,l}^{(i,\,j)} \le 1. \\ 1, & \text{если} \quad l = i. \end{cases} \tag{9}$$

Сформулированная задача является задачей выбора оптимальных потоков и определения оптимальных маршрутов в сети передачи данных по критерию средней задержки [9-15].

При формировании интегрального критерия маршрутизации следует учитывать следующие дополнительные условия и ограничения:

1) для обеспечения заданных значений качества существенное значение имеет значение вероятности потерь пакетов в сети. Потери пакетов в сети могут быть обусловлены целым рядом причин [11-14], но в данной постановке они обусловлены конечными размерами буферов на узлах коммутации пакетов;

2) в МСС могут возникать отказы линий связи;

3) при выборе оптимальных маршрутов необходимо учесть уровень рисков информационной безопасности, которые могут возникнуть в сети передачи данных.

4) целесообразно определить подходы к решению данной задачи, когда потоки $f_{k,l}$ отличаются от пуассоновских.

5) время, через которое в МСС могут произойти изменения ее состояния, требующие решения задачи перераспределения потоков и маршрутов, много больше времени решения этой задачи и реализации управленческих решений.

Метод выбора оптимальных потоков и определения маршрутов в МСС в соответствии с интегральным критерием качества. Задача выбора альтернативных маршрутов относится к классу многопродуктовых задач нелинейного математического программирования с выпуклой целевой функцией и выпуклым множеством ограничений [9-11]. Поэтому ее решение имеет единственный локальный минимум, который также является и глобальным. Для нахождения этого минимума можно применять известные высокоэффективные методы оптимизации. Такими методами, например, могут быть различные градиентные методы, методы девиации (отклонения) потока, а также их различные модификации [12-15].

В основу предлагаемого метода решения задачи положены комбинации методов нечеткой кластеризации, методов нечеткого продукционного логического вывода и методов нечеткого математического программирования. Предлагаемый метод имеет следующий вид.

Шаг 1. На основании данных мониторинга технического состояния МСС, а также на основании мониторинга текущего риска угроз информационной безопасности, в *i*-й момент времени оценивается нечеткая ситуация сетевого состояния для каждого направления (*k, l*) в виде кортежа:

$$\hat{S}^{(k,l)}_{i\,общ} = < \hat{S}^{(k,l)}_{i\,T_{ср}},\ \hat{S}^{(k,l)}_{i\,R},\ \hat{S}^{(k,l)}_{i\,H},\ \hat{S}^{(k,l)}_{i\,P_{loss}} >, \tag{10}$$

где $\hat{S}^{(k,l)}_{i\,T}$ – оценка нечеткой ситуации относительно средней сетевой задержки передачи пакетов; $\hat{S}^{(k,l)}_{i\,R}$ – оценка нечеткой ситуации относительно уровня рисков угроз сетевой информационной безопасности; $\hat{S}^{(k,l)}_{i\,H}$ – оценка нечеткой ситуации относительно состояния технического состояния аппаратно-программных средств и линий связи; $\hat{S}^{(k,l)}_{i\,P}$ – оценка нечеткой ситуации относительно значения вероятности потери пакетов.

Формально правило нечеткого продукционного логического вывода имеет следующий вид:

$$\textbf{ЕСЛИ}\ \hat{S}^{(k,l)}_{i\,T_{ср}} \in \boldsymbol{A}\ \boldsymbol{И}\ \hat{S}^{(k,l)}_{i\,R} \in \boldsymbol{B}\ \boldsymbol{И}\ \hat{S}^{(k,l)}_{i\,H} \in \boldsymbol{C}\ \boldsymbol{И}\ \hat{S}^{(k,l)}_{i\,P_{loss}} \in \boldsymbol{D},\ \textbf{ТО}\ \hat{S}^{(k,l)}_{i\,общ} \in \boldsymbol{U}, \tag{11}$$

где ***A***, ***B***, ***C***, ***D*** и ***U*** – соответствующие области возможных значений соответствующих нечетких ситуаций.

Принятие решения производится в соответствие со следующими правилами.

1.1. ЕСЛИ $\mu^*(\hat{S}^{(k,l)}_{i\,общ}) \geq \mu^*(A_{k,l})\ \forall (k,l) \in N,$ **ТО** «Параметры функционирования сети соответствуют заданным».

РЕШЕНИЕ: «Продолжение сетевого мониторинга, изменений сетевых параметров и сетевой топологии нет».

1.2. ЕСЛИ $\exists (k,l)$ *такие, что* $\mu^*(B_{k,l}) \leq \mu^*(\hat{S}^{(k,l)}_{i\,общ}) < \mu^*(A_{k,l})$, **ТО** «Параметры функционирования сети ухудшились, но ее функционирование возможно».

РЕШЕНИЕ: «Подготовка варианта перераспределения потоков, с учетом возможности исключения линии связи (*k, l*) из состава сети».

1.3. ЕСЛИ $\exists (k,l)$ *такие, что* $\mu^*(\hat{S}^{(k,l)}_{i\,общ}) < \mu^*(B_{k,l})$, **ТО** «Сетевые параметры функционирования ухудшились, ее функционирование невозможно».

РЕШЕНИЕ: «Реализация варианта перераспределения потоков с учетом вывода линии связи (*k, l*) из состава сети».

При этом используются следующие новые обозначения: $\mu^*(\hat{S}^{(k,l)}_{i\,общ})$ – значение функции принадлежности для нечеткой ситуации $\hat{S}^{(k,l)}_{i\,общ}$; $\mu^*(A_{k,l})$, $\mu^*(B_{k,l})$ – пороговые значения функций принадлежности, определяющие области, характеризующие сетевое состояние.

Введем в рассмотрение индикаторную функцию вида:

$$I(k,l) = \begin{cases} 1, & \text{если}\quad \mu^*(\hat{S}^{(k,l)}_{i\,общ}) \geq \mu^*(A_{k,l}), \\ 0, & \text{если}\quad \mu^*(\hat{S}^{(k,l)}_{i\,общ}) < \mu^*(B_{k,l}). \end{cases} \tag{12}$$

ШАГ 2. Проверка сетевой связности. Для этого применяются алгоритмы Дейкстры или Беллмана-Форда.

ШАГ 3. Координация структуры целевого функционала и ограничений с помощью индикаторной функции и минимизация целевого функционала в виде:

$$T_{cp} = \frac{1}{\gamma}\sum_{k=1}^{N}\sum_{l=1}^{N}\lambda_{k\,l}\,t_{k\,l}\,I(k,l) \to min,$$
$$\sum_{k=1}^{N} I(k,l)\,z_{k,l}^{(i,j)} - \sum_{k=1}^{N} I(k,l)\,z_{l,k}^{(i,j)} = \begin{cases} -1, & если \quad l = j, \\ 0, & если \quad l \neq i, j, \\ 1, & если \quad l = i. \end{cases} \qquad (13)$$

ШАГ 4. Проверка выполнения условия 1.1:

4.1 ЕСЛИ условие 1.1 выполняется, **ТО** «Параметры функционирования сети соответствуют заданным, продолжить мониторинг».

4.2 ЕСЛИ условие 1 не выполняется, **ТО** «Запрос ресурса у вышестоящей системы».

4.3 ЕСЛИ «Ресурс получен», **ТО** переход к шагу 2.

Конец.

Таким образом, расширение количества метрик и их атрибутов при решении задачи выбора оптимальных потоков и определения маршрутов в МСС в соответствие с интегральным критерием качества сводится к решению задачи оперативной координации и согласования структуры оптимизируемого функционала и области ограничений, на которой ищется оптимальное решение. Метод и процедуры координации реализуются на основе оперативных методов оценки нечеткой ситуации состояния исследуемой системы, а также методов нечеткой координации и согласования параметров взаимоувязанной иерархической совокупности оптимизационных задач управления МСС.

Отметим, что вероятность потерь P_{loss} для МСС с ограниченным буфером определяется для системы типа *G/G/1/Q* (произвольные вероятностные характеристики входного потока, произвольные вероятностные характеристики обслуживания входного потока на маршрутизаторах, один входной канал, емкость буферного накопителя равна Q), которая является наиболее общей моделью, в соответствие со следующим выражением [9-11]:

$$P_{loss} \approx \frac{1-\rho}{1-\rho^{\frac{2}{K_a^2+K_s^2}Q+1}}\rho^{\frac{2}{K_a^2+K_s^2}Q}, \qquad (14)$$

где ρ – коэффициент использования линии связи (отношение значений интенсивности входящего потока к интенсивности его обслуживания), Q – емкость буфера, K_a – значение коэффициента вариации входного трафика, K_s – значение коэффициента вариации процесса обработки трафика на узле коммутации. Эти значения, а также текущие значения интенсивностей трафика, предлагается оценивать с помощью гибридных адаптивных псевдоградиентных процедур [16-18].

Экспериментальная оценка алгоритмов выбора оптимальных потоков и определения маршрутов в МСС. Для экспериментальной оценки алгоритмов выбора оптимальных потоков и определения маршрутов использовался фрагмент МСС, состоящий из шести узлов, связанных по схеме «звезда». Рассматривались три варианта функционирования данного фрагмента.

При первом варианте все линии связи и все узлы коммутации исправны, рисков угроз информационной безопасности нет. Все потоки имеют пуассоновское распределение.

При втором варианте вследствие отказов либо деструктивных воздействий на МСС имеются отказы некоторых линий связи. Отключение данных каналов связи из процесса передачи по ним информационных потоков также может быть обусловлено высоким уровнем рисков угроз информационной безопасности.

При третьем варианте вероятностные свойства циркулирующих в сети потоков отличны от пуассоновских.

Существует несколько вариантов аналитической аппроксимации зависимости средней временной задержки от параметров потока и характеристик обслуживающего прибора для систем класса *G/G/1* [9, 10]. Одним из наиболее часто применяемых функциональных представлений

является зависимость, в которой верхняя граница суммарного времени задержки обслуживания заявки и её передачи в линию связи имеет следующий вид [10]:

$$t_{k,l} \le \frac{\rho_{k,l}}{f_{k,l}} + \frac{K^2_{a\,k,l} + \rho^2_{k,l} K^2_{s\,k,l}}{2f_{k,l}\,(1-\rho_{k,l})}, \qquad (15)$$

где $t_{k,l}$ – время задержки обслуживания и её передачи в линию связи (k, l); $\rho_{k,l}$ – коэффициент использования линии связи; $K^2_{va\,k,l}$ – значение квадрата коэффициента вариации входящего потока; $K^2_{vb\,k,l}$ – значение квадрата коэффициента вариации процесса обслуживания заявки в линии (k, l).

Отметим, что функционал (15) при подстановке в него значений коэффициентов вариации, равных единице, обладает всеми свойствами пуассоновских потоков. Следовательно, допустимо применение всех методов его оптимизации (минимизации $T_{ср}$), рассмотренных для системы M/M/1 (система с пуассоновским входным потоком и экспоненциальным распределением времени обслуживания поступающих на узел коммутации пакетов, с одним обслуживающим каналом). При этом значения всех параметров трафика и характеристик его обработки можно определить с помощью векторной адаптивной псевдоградиентной процедуры [16-18].

На рисунке 1 приведены сравнительные характеристики зависимостей средней задержки $T_{ср}$ для одного канала передачи данных от коэффициента использования канала ρ. Значение входного потока выбрано равным $\gamma = 500$. Коэффициенты вариации равны: $K_{va\,k,l} = 20$, $K_{s\,k,l} = 1$.

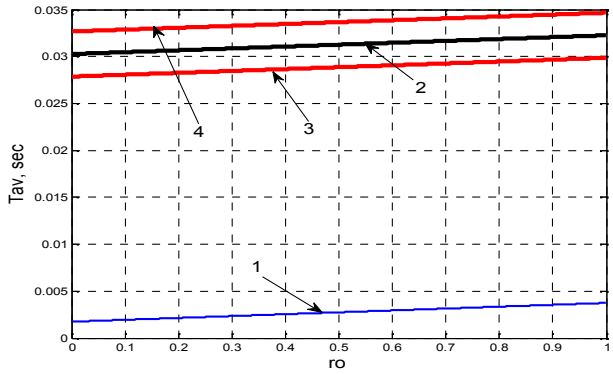


Рис. 1. Зависимости средней задержки от коэффициента использования канала (1 – без учета характеристик входного потока; 2 – истинная характеристика; 3, 4 – зоны возможной абсолютной погрешности оценки)

Кривая 1 построена без учета вероятностных свойств входящего потока. Априори предполагается, что поток является пуассоновским. Кривая 2 соответствует истинной зависимости $T_{ср}$ от ρ с учетом статистических характеристик входного потока. Кривые 3 и 4 определяют зону возможной абсолютной погрешности оценки $T_{ср}$, которые построены на основе данных оценок коэффициентов вариации с помощью адаптивной псевдоградиентной процедуры. Адаптивная псевдоградиентная процедура функционирует в режиме реального времени. Достигаемая средняя относительная погрешность оценки коэффициентов вариации составляет $\delta \le 8\%$, что для систем реального времени является вполне удовлетворительным значением. Характеристики средней сетевой задержки, полученные без учета характеристик входного потока, являются сильно оптимистическими и не соответствующими реальной ситуации, что может привести к деградации сетевых приложений реального времени. Предлагаемый метод позволяет получить абсолютную погрешность оценки средней задержки равную $\Delta \le 8{,}2$ %.

В таблице 1 приведены сравнительные характеристики алгоритмов маршрутизации. Приведены следующие протоколы: Routing Information Protocol (RIP); Border Gateway Protocol (BGP); Open Shortest Path First (OSPF); Internet Gateway Routing Protocol (IGRP) and Enhanced-IGRP (EIGRP); Intermediate System to Intermediate System (IS-IS) protocol.

Таблица 1

Сравнительные характеристики алгоритмов адаптивной маршрутизации

Метрики	Алгоритмы (протоколы) маршрутизации					
	RIP	BGP	OSPF	IGRP/ EIGRP	IS-IS	Интегральный критерий
Длина маршрута	Учет по умолчанию	–	–	–	–	–
Задержка	–	–	–	Учет по умолчанию	Может быть учтена	Учитывается
Надежность	–	–	–	Может быть учтена	Может быть учтена	Учитывается
Полоса пропускания канала связи	–	–	Учет по умолчанию	Учет по умолчанию	–	Учитывается
Нагрузка канала	–	–	–	Может быть учтена	–	Учитывается
Произвольно задаваемый	–	Учет по умолчанию	–	–	Учет по умолчанию	Может быть учтена
Уровень рисков угроз ИБ	–	–	–	–	–	Учитывается
Вероятностные характеристики потока	–	–	–	–	–	Учитывается
Уровень потерь пакетов	–	–	–	Учитывается апостериорно	–	Учитывается априорно

Из таблицы видно, что предлагаемый метод адаптивной маршрутизации на основе интегрального критерия качества позволяет наиболее полно учесть все основные факторы, влияющие на нормальное функционирование МКС. Вычислительная сложность предложенного метода равна $O(M^3)$, где M – число узлов в сети.

Заключение

Результаты анализа численного моделирования предложенных методов и алгоритмов адаптивной альтернативной маршрутизации на основе интегрального критерия качества показали свою высокую эффективность, они достаточно просто реализуются на основе существующих аппаратно-программных платформ. Для их реализации необходимы минимальные вычислительные ресурсы. Применение предложенных алгоритмов позволит сократить время реакции системы управления МСС на возникающие различные изменения сетевого состояния, включая различные внешние деструктивные воздействия на инфраструктуру МСС. Актуальным направлением дальнейших исследований может быть разработка способов оперативного самообучения при различных сетевых ситуациях применяемых интеллектуальных методов управления маршрутизацией в МСС.

Работа проводилась при частичной финансовой поддержке РФФИ (проекты 16-29-09482, 18-07-01369 и 18-07-01488) и бюджетной темы № АААА-А16-116033110102-5.

ЛИТЕРАТУРА

1. Котенко И.В., Саенко И.Б. Архитектура системы интеллектуальных сервисов защиты информации в критически важных инфраструктурах // Труды СПИИРАН. № 1 (24). 2013. С.21-40.
2. Котенко Д.И., Котенко И.В., Саенко И.Б. Методы и средства моделирования распределенных атак в больших компьютерных сетях: состояние проблемы // Труды СПИИРАН. № 3 (22). 2012. С.5-30.
3. Ash G.R. Dynamic Routing in Telecommunications Networks. McGraw–Hill, 1998.
4. Braun H.W. RFC 1104 “Models of Policy Based Routing”.
5. Clark D. D. RFC 1102 “Policy Routing in Internet Protocols”.
6. Comer D. Internetworking with TCP/IP. Volume I: Principles, Protocols and Architecture. New York: Prentice Hall, 2000.

7. Black U. IP Routing Protocols: RIP, OSPF, BGP, PNNI & Cisco Routing Protocols. New York: Prentice Hall, 2000.
8. Armitage G. Quality of Service in IP Networks. New York: Macmillan Technical Publishing, 2000.
9. Клейнрок Л. Теория массового обслуживания. М.: Машиностроение, 1979.
10. Клейнрок Л. Вычислительные системы с очередями. М.: Мир, 1979.
11. Берсекас Д., Галлагер Р. Сети передачи данных. М.: Мир, 1989.
12. Gerla M., Kleinrock L. On the topological design of distributed computer networks // IEEE Trans. on Comm. 1977. Vol. 25. No. 1. Pp. 48–53.
13. Dannis J.E., More J.J. Quasi-Newton methods, motivation and theory // SIAM Review. 1977. Vol. 19. No. 1. P. 46-89.
14. Daniel J.W. The approximate minimization of functional. New York: Prentice Hall, 1971.
15. Fratta L., Gerla M., Kleinrock L. The flow deviation method: An approach to store-and- forward communication network design // Networks. 1973. Vol.3. No. 2. Pp. 97–133.
16. Саенко И.Б., Агеев С.А., Шерстюк Ю.М., Полубелова О.В. Концептуальные основы автоматизации управления защищенными мультисервисными сетями // Проблемы информационной безопасности. Компьютерные системы. №3. 2011. С. 30-39.
17. Саенко И.Б., Агеев С.А., Бушуев А.С., Егоров Ю.П. Концепция автоматизации управления информационной безопасностью в защищенных мультисервисных сетях специального назначения // Автоматизация процессов управления. № 1(23). 2011. С. 50-57.
18. Агеев С. А., Саенко И. Б., Котенко И. В. Метод и алгоритмы обнаружения аномалий в трафике мультисервисных сетей связи, основанные на нечетком логическом выводе // Информационно-управляющие системы. 2018. № 3. С. 61–68.

S.A. Ageev, I.B. Saenko (Saint-Petersburg institute for informatics and automation of Russian Academy of Sciences, Saint-Petersburg)

Operational management of routing in multiservice communication networks taking into account support of information and network security in case of incompleteness of controlled information

The paper offers the new approach to a solution of the problem of the adaptive routing in multiservice communication networks based on combined use of a method of multipath routing of data streams and integral criterion of quality which is created based on the analysis of assessment of an fuzzy hierarchical situation of a current state of a network. The assessment is created by results of permanently the carried-out monitoring of a network by means of methods of fuzzy classification and fuzzy logical inference. Unlike the known methods and protocols of the adaptive routing, in the offered approach the level of information security and also technical condition of network elements and probability of loss of packets in addition are considered.

КОНЦЕПТУАЛЬНЫЕ ОСНОВЫ ПОСТРОЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ АНАЛИТИЧЕСКОЙ ОБРАБОТКИ ЦИФРОВОГО СЕТЕВОГО КОНТЕНТА С ЦЕЛЬЮ ОБНАРУЖЕНИЯ И ПРОТИВОДЕЙСТВИЯ НЕЖЕЛАТЕЛЬНОЙ, СОМНИТЕЛЬНОЙ И ВРЕДОНОСНОЙ ИНФОРМАЦИИ

И. В. Котенко[1], **И. Б. Саенко**[2]
(*Санкт-Петербургский институт информатики и автоматизации
Российской академии наук (СПИИРАН),
199178, С.-Петербург, 14-я линия В.О., 39,
(812) 328-71-81, {ivkote, ibsaen}@comsec.spb.ru*)

Аннотация

В докладе рассматриваются концептуальные основы построения перспективных интеллектуальных систем аналитической обработки цифрового сетевого контента, предназначенных для защиты от нежелательной, сомнительной и вредоносной информации, содержащейся в информационных объектах сети Интернет и социальных сетей. Рассмотрены основные определения, принципы организации, цели и задачи, особенности Интернет-пространства и механизмы обеспечения работы таких систем. Предложена типовая структура интеллектуальной системы аналитической обработки цифрового сетевого контента.

Ключевые слова: защита от информации, интеллектуальная система, аналитическая обработка, цифровой сетевой контент.

Введение

Проблема защиты от нежелательной и вредоносной информации, распространяющейся в Интернете и социальных сетях, является достаточно актуальной для информационной технологии обработки данных [1-6].

В настоящее время эта проблема имеет крайне малое количество научно-технических решений. При этом известные автоматизированные средства выявления и противодействия нежелательной, сомнительной и противоречивой информации в цифровой сетевом контенте не отвечают требованиям по оперативности, полноте, точности и адекватности принимаемых решений [7-9].

Это объясняется, во-первых, сложной иерархической структурой веб-ресурсов, состоящих из множества разнородных элементов. Во-вторых, необходимостью оперативной обработки больших данных, к категории которых относится цифровой сетевой контент. В-третьих, алгоритмы классификации плохо работают на наборах данных малого объема. Наконец, анализ цифрового сетевого контента связан с обработкой противоречивых и изменчивых данных, что определяет необходимость применения методов обработки неполных, противоречивых и нечетких знаний.

Решение этой проблемы связано с разработкой новых моделей, методов и методик, относящихся к таким областям научного знания, как интеллектуальный анализ данных, обработка больших данных, обработка неполных, противоречивых и нечетких знаний и другие.

Все эти модели, методы и методики должны быть реализованы в рамках интеллектуальной системы аналитической обработки цифрового сетевого контента (ИСАОЦСК). Рассмотрение концептуальных основ построения ИСАОЦСК является целью настоящего доклада.

Описание концептуальных основ

Основные определения. Ключевым определением является определение информационного объекта. Понятие *информационного объекта* в общем случае достаточно неопределенно и может включать в себя любой класс сущностей, доступ к которым может нанести ущерб информационной безопасности пользователя [10]. К таковым классам могут относиться исполняемые файлы, Web-сайты или их отдельные страницы, электронные почтовые сообщения, источники

[1] Доктор технических наук, профессор, заведующий лабораторией проблем компьютерной безопасности.
[2] Доктор технических наук, профессор, ведущий научный сотрудник.

информационных контейнеров для различных систем обмена данными (логические узлы пиринговых, почтовых сетей, физические хосты) и т.д.

Принципы организации ИСАОЦСК. Основными требованиями, предъявляемыми к построению и функционированию ИСАОЦСК, являются требования адаптивной и высоко масштабируемой аналитической обработки информационных объектов, обеспечивающей обнаружение нежелательного цифрового контента в реальном или близком к реальному масштабе времени, а также организация эффективного противодействия этому контенту, в случае необходимости. Для реализации этих требований следует использовать методы, модели, методики и алгоритмы, которые ориентированы на следующие аспекты:

1) одновременный анализ большого количества источников данных о смысловом наполнении информационного объекта, в том числе текста, изображений, ссылок, экспертных оценок и т.д.;

2) использование масштабируемой многоуровневой архитектуры системы принятия решений о категориях, к которым относится информационный объект;

3) применение технологии обработки больших данных;

4) применение эффективных алгоритмов переобучения существующих моделей классификаторов и добавления новых без прерывания работы системы анализа;

5) использование различных способов противодействия нежелательной информации, в зависимости от целевой аудитории данной системы.

Учет этих аспектов позволяет сформулировать следующие основные принципы организации ИСАОЦСК:

1) предварительная оценка (категоризация) информационного объекта. Этот принцип обеспечивает возможность формирования начального суждения о принадлежности объекта к той или иной категории и построения окончательного суждения об информационном объекте в случае, если другая информация о нем недоступна;

2) распараллеленный, иерархический анализ различных аспектов данных об объекте в рамках трех основных направлений: а) содержимого объекта, б) его структуры; в) его истории. В общем случае применение группы простых классификаторов дает больший выигрыш в быстродействии, точности и сопровождаемости системы категоризации по сравнению с применением одного сложного;

3) применение методов объединения результатов (предсказаний) отдельных классификаторов для получения окончательного суждения об объекте. В результате, каждый из основных этапов принятия решения о категории информационного объекта может быть привязан к отдельному аспекту его состояния;

4) использование различных способов противодействия нежелательной информации. Так, например, для защиты от рекламных почтовых сообщений, оптимальным способом является автоматическое удаление, для защиты детей от нежелательной для них информации – автоматическая переадресация, для информации, запрещенной законодательством Российской Федерации, – блокирование, и т.д.

Цели и задачи ИСАОЦСК. Основными целями функционирования ИСАОЦСК являются выявление в информационных объектах сети Интернет и в социальных сетях нежелательной, сомнительной и вредоносной информации и выработка мер противодействия эти видам информации.

Основными задачами, решаемыми ИСАОЦСК, являются:

1) получение информации из множества гетерогенных источников о возможных источниках хранения и распространения нежелательной, сомнительной и вредоносной информации;

2) определение тематической направленности анализируемой информации;

3) противодействие загрузке, отображению и распространению нежелательной, сомнительной и вредоносной информации;

4) отображение получаемых данных и результатов их обработки в удобном графическом интерфейсе;

5) выявление путей распространения нежелательной, сомнительной и вредоносной информации в социальных сетях.

Особенности Интернет-пространства и социальных сетей, влияющие на построение ИСАОЦСК. К числу основных особенностей сети Интернет и социальных сетей, влияющих на построение ИСАОЦСК, относятся следующие факторы:

1) огромный объем данных, доступных пользователям, обуславливает высокие требования к свойству оперативности, налагаемые на разрабатываемые решения. Суммарный объем данных из года в год растет, что обуславливает необходимость в поиске легко масштабируемых решений;

2) типичный информационный объект имеет сложную, многокомпонентную структуру, что влечет необходимость применения методов анализа различных типов данных и объединения его результатов;

3) современное информационное пространство также характеризуются большим количеством связей между отдельными объектами, что также должно учитываться при анализе контента;

4) при прямом запрете доступа к информационному объекту, пользователь скорее всего будет предпринимать попытки обойти запрет, поэтому важной задачей является переключение его внимания на легитимные информационные объекты.

Механизмы обеспечения работы ИСАОЦСК. К числу основных механизмов, которые предлагается включить в состав ИСАОЦСК, но которые отсутствуют у известных систем такого типа, например, у автоматизированных систем Интернет-контента, относятся следующие механизмы:

1) распределенного интеллектуального сканирования сетевого цифрового контента, обеспечивающие оперативный, надежный и устойчивый к внешним воздействиям сбор и предварительную оценку информационных ресурсов сети Интернет и социальных сетей;

2) многоаспектной оценки и категоризации смыслового наполнения информационных объектов, содержащие средства анализа текстовых, структурных, адресных, связностных, мультимедийных и других признаков;

3) обеспечения своевременности многоуровневого и многомодульного анализа информационных объектов на основе использования параллельных вычислений;

4) устранения неполноты и противоречивости оценки и категоризации смыслового наполнения информационных объектов на основе использования методов обработки неполных, противоречивых и нечетких знаний;

5) адаптации и переобучения системы анализа информационных объектов, в том числе в режиме эксплуатации;

6) выработки и выбора мер противодействия нежелательной, сомнительной и вредоносной информации;

7) реализации визуальных интерфейсов для выявления и противодействия нежелательной, сомнительной и вредоносной информации.

Распределенные интеллектуальные сканеры сетевого цифрового контента должны обеспечивать децентрализованные режимы сбора и предобработки данных, содержащихся в информационных объектах, за счет наделения сканеров интеллектуальными свойствами. Основными функциями предварительной обработки данных являются фильтрация, нормализация (приведение к общему формату), обобщение и приоритезация данных [11]. Особенностью функционирования интеллектуальных сканеров является повышение оперативности сбора данных за счет применения интеллектуальных методов распределенной потоковой обработки информации.

Механизмы многоаспектной оценки и категоризации смыслового наполнения информационных объектов должны использовать методы интеллектуального анализа данных и основанные на них классификаторы, которые учитывают достаточно большое количество признаков [12]. К их числу относятся текстовые, структурные, адресные, связностные, мультимедийные и другие признаки, а также признаки, извлекаемые из внешних источников данных об информационном объекте (whois-серверы, списки классифицированных объектов, репутационные системы и т.д.). Для анализа информационных объектов по извлеченным и сформированным признакам применяется многоуровневая классификация [13, 14]. Элементы первого, начального уровня представляют собой функциональные блоки, ориентированные на отдельные категории. Они используют предварительно обученные классификаторы, дающие суждения о том, отно-

сится ли тот или иной вектор, характеризующий конкретный информационный объект, к определенной категории (таким образом, элементы первого уровня решают задачу бинарной классификации). Элементы второго уровня используют классификаторы, ориентированные на принятие решения о принадлежности вектора описаний заданного информационного объекта одной из заданных тем (категорий) в рамках информации, получаемой при анализе отдельных аспектов объекта (например, текстовое содержимое, структурные особенности, источник и т.д.). Результаты классификации, полученные на втором уровне, используются для формирования описаний анализируемых информационных объектов, которые используются для обучения и принятия решений элементом третьего уровня, ориентирующегося на формирование окончательных решений.

Механизмы обеспечения своевременности многоуровневого и многомодульного анализа информационных объектов предлагается реализовать на технологии параллельной обработки данных. В качестве прототипа для разработки такой платформы рассматриваются системы Hadoop, Spark, Flink и другие. В качестве основных показателей, используемых для выработки требований и оценки функционирования этих компонентов, предлагается учитывать показатели масштабируемости и надежности хранения информации. Требования по масштабируемости удовлетворяются за счет изменения количества исполнительных вычислительных узлов и процессов в платформе. Требования надежности реализуются за счет организации и использования для хранения информации распределенной файловой системы Hadoop Distributed File System (HDFS).

Механизмы устранения неполноты и противоречивости оценки и категоризации смыслового наполнения предназначены усовершенствование классификаторов информационных объектов с целью придания им способности обработки неполных, противоречивых и нечетких знаний. Неполнота и противоречивость оценки и категоризации смыслового наполнения информационных объектов свойственны этому процессу в силу самого характера распространения вредоносной и нежелательной информации в сети Интернет и в социальных сетях. Для достижения этой цели предлагается использовать методы нечеткого оценивания ситуаций, нечеткого вывода и нечеткой оптимизации. Эти методы в предметной области аналитической обработки цифрового контента будут применяться одними из первых.

Механизмы адаптации и переобучения системы анализа информационных объектов, в том числе в режиме эксплуатации должны использовать экспертные знания и существующие списки классифицированных информационных объектов. Применяемые методы выделения признаков должны позволять учитывать при переобучении максимально разнообразные аспекты информационных объектов, в том числе историю изменений и данные, поступающие из сторонних источников. Компоненты должны иметь открытую масштабируемую архитектуру, что позволит добавлять и адаптировать в системе новые способы обработки цифрового контента без существенных затрат на переобучение.

Механизмы выработки и выбора мер противодействия нежелательной, сомнительной и вредоносной информации отличаются от существующих аналогов использованием особенностей хранилища информации, построенного на основе файловой системы HDFS, а также применением механизма логического вывода, основанного на экспертных знаниях, использованием различных методик вычисления показателей (метрик) вредоносности и опасности цифрового контента и анализа рисков для последующего выбора соответствующих контрмер, а также сочетанием как методов автоматического вывода, так и визуального анализа результатов анализа для формирования мер противодействия.

Механизмы реализации визуальных интерфейсов для выявления и противодействия нежелательной, сомнительной и вредоносной информации отличаются от известных тем, что они, как правило, ориентированы на использование нестандартных моделей визуализации, к числу которых следует отнести такие модели, как графы с глифами, карты деревьев, диаграммы Вороного и другие, а также модели виртуальной и дополненной реальности.

Структура ИСАОЦСК. Результаты выявления механизмов обеспечения работы ИСАОЦСК позволяют предложить типовую структуру системы такого типа, показанную на рисунке 1.

В типовой структуре ИСАОЦСК выделяются следующие уровни:

1) уровень сбора и предварительной обработки данных о безопасности сетевых информационных объектов;

2) уровень оценивания смыслового содержания информации;
3) уровень выработки мер противодействия.



Рис. 1. Типовая структура ИСАОЦСК

Исходными данными для ИСАОЦСК являются информационные объекты сети Интернет и социальных сетей.

Результаты, полученные с помощью предлагаемой системы, используются администраторами безопасности, отвечающими за защиту от нежелательной, сомнительной и вредоносной информации. Потребителями результатов функционирования системы являются организации телекоммуникационного сектора, силовые министерства и ведомства, административно-государственные структуры, в которых предъявляются повышенные требования по защите от нежелательной и вредоносной информации.

Заключение

Интеллектуальные системы аналитической обработки цифрового сетевого контента, концептуальные основы построения которых рассмотрены в докладе, предназначены для защиты от сомнительной, нежелательной и противоправной информации. Эта разновидность информации представляет собой достаточно серьезную социальную угрозу и может являться средой распространения призывов к свершению актов терроризма и идеологического экстремизма, а также наносить вред психическому и физическому здоровью граждан и побуждать их к совершению противоправных действий. Тем самым, эта информация представляет собой достаточно существенный источник опасности для общества, экономики и государства.

Предлагаемая типовая структура ИСАОЦСК включает ряд механизмов, которые позволяют выделять эту систему в отдельный класс интеллектуальных систем и отсутствуют у известных автоматизированных систем анализа Интернет-контента. Ориентация этих механизмов на использование интеллектуальных методов, как предполагается, позволяет успешно достичь поставленных перед ИСАОЦСК целей.

Работа проводилась при финансовой поддержке Российского научного фонда (проект 18-11-00302) в СПИИРАН.

ЛИТЕРАТУРА

1. Базенков Н.И., Губанов Д.А. Обзор информационных систем анализа социальных сетей // Управление большими системами. Выпуск 41. М.: ИПУ РАН, 2013. С. 357-394.
2. Котенко И.В., Чечулин А.А., Комашинский Д.В. Автоматизированное категорирование веб-сайтов для блокировки веб-страниц с неприемлемым содержимым // Проблемы информационной безопасности. Компьютерные системы. № 2. 2015. С. 69-79.
3. Котенко И.В., Юсупов Р.М. Информационные технологии для борьбы с терроризмом // Защита информации. INCIDE. № 2. 2009. С. 74-79.
4. Marcus S.E., Moy M., Coffman T. Social Network Analysis / Cook D.J., Holder L.B. (eds.) // Mining Graph Data. John Wisley & Sons, Hoboken, 2007.
5. Scott J. Social network analysis: developments, advances, and prospects // Social Network Analysis and Mining. January 2011. Vol. 1. No. 1. Pp. 21–26.
6. Qi X., Davison B.D. Web Page Classification: Features and algorithms // ACM Computing Surveys (CSUR). 2009.
7. Belmouhcine A., Idrissi A., Benkhalifa M. Web classification approach using reduced vector representation model based on html tags // Journal of Theoretical and Applied Information Technology. 2013. Vol. 55. No. 1. Pp. 137-148.
8. Borgatti S. P., Everett M. G., Freeman L. C. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies, 2002.
9. Calado P., Cristo M., Moura E., Ziviani N., Ribeiro-Neto B., Goncalves M. A. Combining link-based and content-based methods for web document classification // CIKM ’03: Proceedings of the 12th International Conference on Information and Knowledge Managementю New York, NY, ACM Press. 2003. Pp. 394–401.
10. Котенко И.В., Саенко И.Б., Чечулин А.А. Защита от нежелательной и вредоносной информации в глобальных информационных сетях // Информационно-психологическая и когнитивная безопасность. Коллективная монография / Под ред. И. Ф. Кефели, Р. М. Юсупова. СПб.: Изд-во «Петрополис», 2017. С. 175-194.
11. Chakrabarti S., Dom B., Agrawal R., Raghavan P. Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies // VLDB Journal. August 1998.
12. Новожилов Д.А., Чечулин А.А., Котенко И.В. Улучшение категорирования веб-сайтов для блокировки неприемлемого содержимого на основе анализа статистики html-тэгов // Информационно-управляющие системы. № 6. 2016. С. 65-73.
13. Dumais S., Chen H. Hierarchical classification of web content // Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '00). ACM, New York, NY, USA. 2000. Pp. 256-263.
14. Kotenko I., Chechulin A., Komashinsky D. Categorisation of web pages for protection against inappropriate content in the internet // International Journal of Internet Protocol Technology (IJIPT). 2017. Vol. 10. No. 1. Pp. 61-71.

I.V. Kotenko, I.B. Saenko (Saint-Petersburg institute for informatics and automation of Russian Academy of Sciences, Saint-Petersburg)

Common architecture of the intelligent system for analytical processing of digital network content with the purpose of detecting from undesirable information

The paper considers the conceptual foundations of creating the perspective intelligent systems of analytical processing of digital network content intended for protection against the unwanted, doubtful and harmful information contained in information objects of the Internet and social networks. The main concepts, the organization principles, the purpose and the tasks, feature of Internet space and mechanisms of supporting such systems are considered. The typical structure of the intelligent system of analytical processing the digital network content is offered.

ПОДХОД К АНАЛИЗУ ЗАЩИЩЕННОСТИ БЕСПРОВОДНОЙ САМООРГАНИЗУЮЩЕЙСЯ СЕТИ КРИЗИСНОГО УПРАВЛЕНИЯ

В. А. Десницкий[1], **Н. Н. Рудавин**[2]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук,
С.-Петербург, 14-я линия Васильевского острова, 39,
desnitsky@comsec.spb.ru, nikolay-rudavin@yandex.ru*)

Аннотация

Цель работы состоит в решении проблемы анализа безопасности самоорганизующейся сети кризисного управления путем выявления уязвимостей, построения моделей угроз, анализа сценария типовой атаки и актуализации наиболее значимых угроз. Объектом исследования выступает сеть кризисного управления, включающая средства сканирования и дроны, доставляющие аварийных модулей сети для её восстановления. Разработан прототип системы, моделирующий сеть кризисного управления и механизм обеспечения ее доступности. Прототип используется при анализе защищенности сети и проведении экспериментов.

Ключевые слова: беспроводная самоорганизующаяся сеть, кризисное управление, ZigBee, дроны.

Введение

В современном мире благодаря прогрессу в области компьютерных технологий появилась возможность автоматического управления критически важными объектами, где негативное влияние человеческого фактора недопустимо. Таким образом повышается надежность работы систем, где возникновение неисправностей в работе может оказаться фатальным. Для этого создаются сети кризисного управления, включающих комплекс мер и средств, направленных на выход из критических состояний. В случае возникновения нештатной ситуации в такой самоорганизующейся сети, требуется выяснить источник проблемы, причину и с минимальным вовлечением обслуживающего персонала принять решение по устранению проблемы. С ростом развития технических средств передачи данных и самоорганизующихся сетей, применяемых в том числе в кризисном управлении, также растет интерес злоумышленников, ставящих перед собой цель нарушить безопасность систем. Особенно актуальна данная проблема для сетей кризисного управления, так как в них поддержание стабильности работы является приоритетом.

В работе представлена анализируемая киберфизическая система, построенная на базе самоорганизующейся сети, использующей протокол ZigBee. Важной её особенностью является наличие координирующего элемента, на котором установлено дополнительное ПО, отвечающее за стабильную работу системы и оперативное решение задач обеспечения доступности. Элементы сети могут быть стационарные или подвижные, поэтому топология сети может меняться.

При анализе защищенности были составлены модели угроз с применением нескольких классификаторов. Исходя из полученных моделей, выделялись наиболее актуальные угрозы. Показаны примеры сценариев атак по актуальным угрозам с их практической реализацией и получением результатов. Исходя из этого, могут быть предложены методы борьбы с этими угрозами, которые в дальнейшем могут быть применены не только для конкретно спроектированной системы, но и для сетей кризисного управления в целом.

Анализ защищенности

Концепция системы обеспечения доступности. Имеются следующие элементы:

1. средства автоматического сканирования узлов сети и формирования графовой модели сети;
2. алгоритмическое обеспечение анализа динамически изменяемых маршрутов, характерных для самоорганизующихся сетей [1] и поступающих от узлов данных для решения задачи мониторинга связности сети, выявления нарушений доступности узлов в том числе потенциальных;
3. средства принятия решений по автоматизированному пространственному перемещению беспилотных летательных аппаратов (БПЛА), снабженных дополнительными узлами-роутерами

[1] Кандидат технических наук, старший научный сотрудник лаборатории проблем компьютерной безопасности.
[2] Программист лаборатории проблем компьютерной безопасности.

в места реальных и потенциальных нарушений доступности сети для устранения этих нарушений. Возможно зависание дрона на непродолжительное время в области потери сигнала либо выгрузка модуля в заданной точке для восстановления нарушений в структуре сети, повышения пропускной способности и снижения вероятности нарушений их связности в дальнейшей работе.

В рамках предлагаемого подхода используются следующие ключевые элементы: координирующая ЭВМ с беспроводным XBee-модулем в роли координатора; узел-роутер/конечное устройство с автономным источником питания, GPS-приемником и модулем управления (на основе микросхемы Arduino UNO) [2]; аварийный узел-роутер, закрепленный на БПЛА (Parrot AR Drone 2.0). Штатные узлы могут находятся в некоторой стационарной точке, а также на подвижном носителе или транспортном средстве.

Виды угроз системы. В рамках анализа защищенности беспроводной коммуникационной сети (БКС) рассматриваются следующие направления осуществления атакующих воздействий, характерные для самоорганизующихся сетей:

1. пассивная атака на протокол маршрутизации – прослушивание пакетов, посылаемых в соответствии с протоколом маршрутизации;
2. перенаправление пакетов – использование протокола маршрутизации для перенаправления пакетов, идущих от или к целевому узлу, через определённый узел;
3. переполнение таблицы маршрутизации – переполнение таблицы маршрутизации протокола, блокирующее создание новых маршрутов;
4. программный отказ/отключение элемента – узел не предоставляет сервисы сети;
5. отбор информационных ресурсов – атакующий или атакуемый узел склонен использовать общие ресурсы сети больше остальных, в том числе среду передачу данных.
6. отбор энергетических ресурсов (атака истощения энергоресурсов) – нарушитель вынуждает узел производить дополнительные действия, при этом расходуется запасенная энергия целевого узла;
7. определение нарушителем топологии сети – получение информации о сети и расположении ее элементов;
8. спуфинг-атака – атакующий узел выдает себя за некоторый узел сети с определенной ролью, в том числе ролью координатора, получая тем самым возможность производить множество специфичных действий;
9. намеренные помехи (jamming-атака) – канал передачи данных заполняется посторонними шумами, которые могут иметь естественную и искусственную природу, нарушая передачу полезных данных;
10. аппаратный отказ/отключение/повреждение элемента – имея физический доступ к системе или её части, нарушитель способен осуществить повреждение или уничтожение элементов сети.

Классификация угроз. Модель угроз представлена в таблице 1. Приведена классификация по атакуемому элементу системы: сети ZigBee – сети узлов XBee, работающих в режиме передачи API-пакетов (1.1); аппаратной платформе Arduino с электропитанием и датчиками GPS (1.2); координирующей ЭВМ – электронно-вычислительному устройству, производящему определение состояния ZigBee-сети с помощью модуля-координатора XBee, которое принимает решение о восстановлении сети и программирует БПЛА-носители (1.3); БПЛА-носителю AR-Drone, обеспечивающему доставку узлов сети (1.4).

По источнику угроз строится следующая классификация: внешний – источник угроз изначально не имеет прямого отношения к работе сети (либо не взаимодействует с системой по её протоколам, либо получает доступ со временем) (2.1); внутренний, где источник находится внутри системы, и имеется возможность взаимодействовать с одним элементом, либо группой элементов, либо всей системой посредством штатных протоколов (2.2).

Также строится классификация по атакуемому обеспечению: программное (3.1) и аппаратное (3.2).

По характеру умысла угрозы могут быть случайны (4.1) и умышленны (4.2). Случайная – угроза возникает вследствие непреднамеренных событий, может быть как объективной: отказы, сбои, стихийные бедствия, чрезвычайные ситуации, электромагнитная несовместимость, так и субъективной. Умышленная – угроза целенаправленна, атакующий мотивирован и настроен на выполнение задачи нарушения безопасности.

Нарушителем могут преследоваться следующие цели: просмотр – нарушитель хочет получить информацию, но он не заинтересован во вмешательстве в работу системы (5.1); изменение – выполнение системой задач в интересах нарушителя (5.2); уничтожение – не позволить системе выполнять свои основные задачи (5.3).

Реализация атак приводит в следующим последствиям для системы: нарушение конфиденциальности – информация, передаваемая в системе, становится известной субъекту, не располагающему полномочиями доступа к ней (6.1);нарушение доступности – создание условий, при которых доступ к услуге или информации будет заблокирован или возможен за время, которого недостаточно для выполнения задачи (6.2); нарушение целостности – повреждение, модификация или уничтожение информации, хранящейся/передающейся внутри сети или взаимодействующей с внешней средой (6.3).

Таблица 1

Модель угроз системы

Угроза	Подсистемы	Источник	Обеспечение	Умысел	Цель	Последствия
1	1.1, 1.4	2.1, 2.2	3.1	4.2	5.1	6.1
2	1.1, 1.3, 1.4	2.1, 2.2	3.1, 3.2	4.1, 4.2	5.1, 5.2	6.1, 6.2, 6.3
3	1.1	2.1, 2.2	3.1	4.2	5.3	6.2
4	1.1, 1.2, 1.3, 1.4	2.1, 2.2	3.1, 3.2	4.1, 4.2	5.3	6.2, 6.3
5	1.1, 1.2	2.1, 2.2	3.1, 3.2	4.1, 4.2	5.1, 5.3	6.2
6	1.1, 1.4	2.1, 2.2	3.1, 3.2	4.2	5.2, 5.3	6.2, 6.3
7	1.1, 1.2, 1.3, 1.4	2.1, 2.2	3.1, 3.2	4.2	5.1	6.1
8	1.1, 1.2, 1.3, 1.4	2.1	3.1	4.2	5.1, 5.2, 5.3	6.1, 6.2, 6.3
9	1.1, 1.4	2.1	3.2	4.1, 4.2	5.2, 5.3	6.2, 6.3
10	1.1, 1.2, 1.4	2.1, 2.2	3.2	4.1, 4.2	5.2, 5.3	6.2, 6.3
11	1.2, 1.4	2.1	3.1, 3.2	4.2	5.1, 5.2, 5.3	6.1, 6.3

Далее рассмотрены примеры реализации атак на каждый элемент узла. Все приведенные атаки реализуются внешним нарушителем, а их целью, как и следствием является нарушение доступности сети.

Сеть X-Bee – спуфинг-атака. При внедрении атакующего модуля в сеть он может начать реализовывать атаку отбора информационных ресурсов путем регулярной отправки пакетов на другие модули, затрудняя передачу пакетов вплоть до отказа в обслуживании.

Атакующий имеет информацию об идентификаторе сети, а также радио-модуль, поддерживающий протокол ZigBee, подключает своё устройство, в котором уже настроен идентификатор сети PAN_ID, таким образом становясь полноправным узлом сети. Также возможен сценарий подмены, когда нарушитель выводит из строя узел с заранее известным MAC-адресом, и используя модифицированный модуль и клонируя адрес жертвы, подменяет его собой. Подключившись к сети, устройство инициирует replay-атаку, отправляя множество раз в сеть AT-команды, вынуждающие другие устройства генерировать пакеты и отправлять ответы [3]. В случае отправки команды NetworkDiscovery (ND) идет широковещательная отправка пакета, и все узлы сети должны ответить информацией о себе. Учитывая, что не все устройства связаны напрямую, а ответ передать нужно всем узлам, возможна задержка в передачи пакетов, в том числе критически важных для работы системы. Проведен эксперимент с загрузкой сети сообщениями от атакующего устройства для проверки работы передачи пакетов системы (анализ сети координатором и отправка им же широковещательных сообщений).

В рамках эксперимента была поднята сеть из 5 модулей XBee, а также написана программа координирующего для устройства [4]. Производились стандартные поиск и оценка состояния сети с интервалом в 10 секунд. Далее атакуемый модуль начал производить отправку в сеть AT-команд с периодичностью около 10 команд в секунду. В результате из-за возникшей путаницы пакетов координатору перестала своевременно и как следствие корректно приходить часть сообщений алгоритма поиска, и как следствие он стал получать недостоверные данные о состоянии сети. Также была замечена задержка в передаче сообщений практически до секунды. В результате перегрузки сети информация от некоторых устройств оказывалась потерянной, что может пагубно отразиться на оценке состоянии сети и оперативности реагирования на неполадки.

Модуль узла сети - отбор энергетических ресурсов. Подключившись к открытым элементам интерфейса платформы, возможно инициирование лишних операций в аппаратной плат-

форме либо неоправданное увеличение мощности антенны модуля XBee AT-командами, а также переключение модуля из спящего режима в режим постоянной работы. Таким образом, возможно многократное увеличение скорости разряда аккумуляторной батареи атакуемого узла.

Нарушитель имеет физический доступ к устройству и каналу передачи данных между платформой и узлом XBee. Далее, подача модулю AT-команды PowerLevel 4 увеличивает мощность антенны и выводит модуль из спящего режима путем инициирования отправки пакета даже с неверной адресацией, увеличивая расход энергии с 33 мкВт до 165 мВт. Также аппаратной платформе Arduino на входной пин подаются случайные данные, вынуждая устрйоство выходить из спящего режима, увеличивая расход с 40 до 215 мВт (расчет указан для платформы Arduino Uno R3 на чипе 16u2). Таким образом, атакующий воздействует на узел, даже не входя в сеть XBee. Атакующий действует локально, истощая энергоресурсы, и на работе сети изменения узла могут и не отобразиться.

При исследовании угроз был проведен эксперимент с локальным узлом сети, причем использовалось следующее оборудование: микроконтроллер платформы Arduino Uno R3 на чипе 16u2, источник напряжения 5В, амперметр для замера потребления тока, модуль XBee s2 в роли конечного устройства и GPS-модуля NEO-6m.

Контроллер был переведен в режим сна с интервалом пробуждения 5 секунд до момента получения данных. С таким же периодом сна работал модуль XBee, в результате удалось сократить потребляемый ток до 0,03-0,04А. Модулируя атаку, была отправлена AT-команда SP, вернувшая модуль в активный режим, сразу увеличив потребляемый ток до 0,07 А. При этом непрерывном отправлении данных на Arduino и невозможности контроллера «уснуть», потребление тока возросло до 0,09А. В общем энергопотребление возросло в 2 раза, что для труднодоступных узлов сети может оказаться критическим [5].

Координирующая ЭВМ - программный отказ/отключение элемента. При получении доступа к координирующей ЭВМ у нарушителя открывается доступ к XBee модулю-координатору. В этом случае возможно отключение любого из элементов, начиная от конечных устройств и заканчивая самим координатором, отключение которого выводит из строя всю сеть. Нарушитель может получить всю информацию об узлах сети, а также выполнить удаленно манипуляции с ними путем отправки API-пактов с AT-командами. Имея информацию об узлах и топологии сети, нарушитель отправляет на заданный атакуемый узел AT-команду PowerLevel 00. Таким образом, назначив минимальный уровень сигнала антенны, связь с устройством пропадает полностью, а восстановить её можно только подключившись к потерянному модулю локально, дав для этого команду усилить сигнал, или отправить аварийный узел для временного восстановления связи с пропавшим узлом.

В эксперименте была поднята сеть из Координатора (COORDINATOR), и двух роутеров (R1), (R2). Элементы сети были разнесены на расстояния, при которых работа была бы возможна только на максимальном значении уровня мощности антенн [6].

Сеть состоит из 3 элементов, при этом фактическая связь присутствуют только между COORDINATOR и R1 с уровнем сигнала LQI 60,78% и R1 и R2 с уровнем LQI 37,25%. Получив доступ к приложению, роутеру R1 отправляется команда PL 0x00. Снизив мощность антенны до минимума, связь с R1и R2 пропадает, это видно в таблицах программы координирующей ЭВМ (рис. 1). Восстановить связь удаленно уже невозможно.

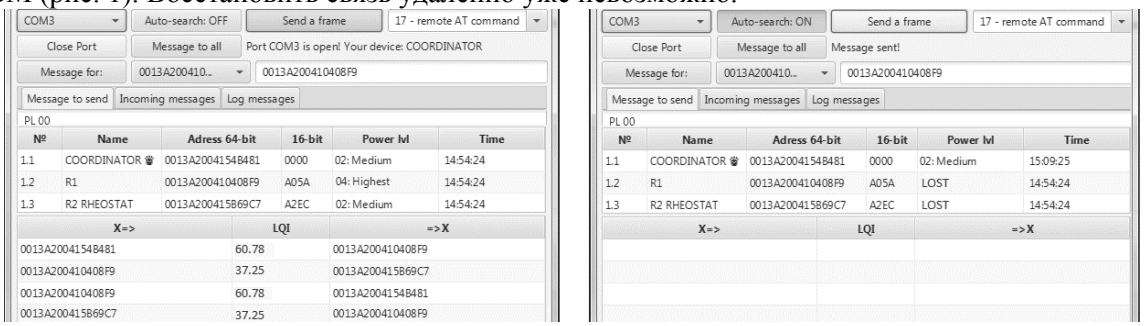


Рис. 1. Состояние сети до атаки (слева) и после (справа)

БПЛА-носитель AR-Drone – помехи. Поскольку дрон периодически использует сеть XBee для получения команд управления автопилоту, в условиях высокой помехообстановки возмож-

ны потери команд управления, к тому же растет энергопотребление модуля управления за счет вынужденного увеличения мощности антенного устройства. Более того система управления получает информацию о местоположении через датчик GPS, расположенный на дроне, и при невозможности получить корректные координаты местоположения БПЛА становится он оказывается неприменим, в том числе невозможен возврат на точку старта. Приведем пример следующих двух сценариев атаки с применением помех.

1. Атака на GPS канал: нарушитель является – постановщик помех и производит атаку с целью разрушения навигационных сообщений. Потеря сигнала приводит к прекращению движения БПЛА по маршруту. Более того, при полете на большой высоте или в условиях ограниченной видимости возможен уход БПЛА с точки его текущего месторасположения, что приведет к его, возможно, безвозвратной потере.

2.Атака на XBee-модуль БПЛА: потеряв сигнал, автопилот продолжает выполнять маршрут, который мог потерять актуальность. При этом, не получив команду с возвратом, аппарат может зависнуть на ранее введенном месте назначения до полного разряда аккумулятора.

Заключение

Результатом работы является составление полной таблицы угроз для БКС кризисного управления. Выделены наиболее актуальные угрозы, связанные со спецификой работы подобных сенсорных сетей [7]. Также на собранном прототипе БКС, применяющей БПЛА удалось смоделировать ряд атак, возможных за счет актуальных угроз, проанализированных ранее. В дальнейшем планируется расширение экспериментов с моделированием атак при большем масштабе БКС, а также разработка концепции защиты сетей кризисного управления с учетом более проработанной практической реализацией защиты.

Работа выполнена при финансовой поддержке Гранта Президента Российской Федерации № МК-5848.2018.9.

ЛИТЕРАТУРА

1. Portmann, M., Pirzada, A.A. Wireless mesh networks for public safety and crisis management applications. In:" in IEEE Engineering Management Review, vol. 39, no. 4, 2011. P. 114-122.
2. Kale, V. S., Kulkarni, R. D. Real Time Remote Temperature & Humidity Monitoring Using Arduino and XBee S2 //International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering. – 2016. – Т. 4. – №. 6.
3. Supporting ZDOs with the XBee API // URL: http://ftp1.digi.com/support/images/APP_NOTE_XBee_ZigBee _Device_Profile.pdf (дата обращения: 06.07.2018).
4. XBee Java Library // URL: https://www.digi.com/resources/documentation/digidocs/90001438/ (дата обращения: 06.07.2018).
5. Десницкий В.А., Котенко И.В. Проектирование защищенных встроенных устройств на основе конфигурирования. Проблемы информационной безопасности. Компьютерные системы, 2013. № 1. С. 44-54.
6. Чечулин А.А., Десницкий В.А., Котенко И.В. Анализ информационных потоков для построения защищенных систем со встроенными устройствами. Системы высокой доступности, 2012. Т. 8. № 2. С. 116-122.
7. Desnitsky, V.A., Kotenko, I.V., Nogin, S.B. Detection of anomalies in data for monitoring of security components in the Internet of Things // Proceedings of International Conference on Soft Computing and Measurements, SCM 2015, 2015. P. 189-192.

V.A. Desnitsky, N.N. Rudavin (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg)

Approach to security analysis of wireless self-organizing network for crisis management

The goal of the work is to tackle a problem of security analysis of mesh networks for crisis management by revelation of vulnerabilities, construction of the threat model, an analysis of a typical attack and determination of the most significant threats. The object of the study is a case of crisis management, including scanning tools and drones delivering emergency modules of the network for its recovery. We developed a prototype of the system, which simulates a crisis management network and a mechanism for ensuring its availability. The prototype is applied by analyzing network security and conducting experiments.

ИДЕНТИФИКАЦИЯ НАИБОЛЕЕ ВЕРОЯТНЫХ ТРАЕКТОРИЙ СОЦИОИНЖЕНЕРНЫХ АТАК В УПРАВЛЕНИИ РИСКАМИ, АССОЦИИРОВАННЫМИ С ПОЛЬЗОВАТЕЛЯМИ

А. О. Хлобыстова[1], **М. В. Абрамов**[2], **А. Л. Тулупьев**[3]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук,
199178, С.-Петербург, 14-я лин. В.О., 39;
Санкт-Петербургский государственный университет,
198504, С.-Петербург, Петергоф, Университетский пр., 28;
akhlobystova@yandex.ru, mva16@list.ru, alexander.tulupyev@gmail.com*)

Аннотация

Цель настоящей работы заключается в разработке подходов к решению задачи выявления наиболее критичных траекторий распространения многоходовых социоинженерных атак и их последующая реализация. Формально задача сводится к нахождению в графе пути, в котором произведение весов всех рёбер, входящих в данный путь, максимально. Трудности, возникающие при более детальном рассмотрении поставленной задачи, заключаются в том, что граф является полным графом с многочисленным числом вершин, что делает процесс ресурсозатратным, а также невозможно применение известных алгоритмов по нахождению путей в графе в связи с тем, что веса рёбер — это вероятности. Последняя упомянутая проблема может быть решена путём рассмотрения величин обратных к значениям весов рёбер, что позволяет перейти к поиску минимального пути в графе, также возможно логарифмирование весов, вследствие чего вместо произведения весов будет рассмотрена их сумма. Таким образом, задача сводится к стандартному поиску наикратчайшего пути в графе. Плотность графа предлагается уменьшить путём исключения из рассмотрения дуг с малыми вероятностями. Основным результатом работы является нахождение оптимального алгоритма поиска наиболее уязвимых для атаки траекторий и его программная реализация.

Ключевые слова: социоинженерные атаки, защита пользователя, информационная безопасность, анализ защищённости, траектории атаки.

Введение

Сегодня осуществляется активная деятельность злоумышленников, использующих не только технические средства для получения доступа к конфиденциальным данным, но и прибегающих к методам социальной инженерии. Согласно [1, 2], рост числа киберпреступлений интенсифицируется с каждым годом всё сильнее. Вместе с тем растёт и число социоинженерных атак на пользователей информационных систем. Таким образом актуальной видится задача повышения уровня защищённости пользователей информационных систем от социоинженерных атак. Важной составляющей этой задачи является задача анализа защищённости пользователей информационных систем от социоинженерных атак. Решение таких задач играет существенную роль в управлении рисками, ассоциированными с пользователями.

Социоинженерные атаки могут быть одноходовые (прямые) и многоходовые, совершаемые через цепочку пользователей, где начальный и конечный пользователи не совпадают. Подходы к оценке защищённости пользователей от прямых социоинженерных атак наиболее подробно изложены в [3]. В данной статье исследуется вопрос управления рисками, ассоциированными с персоналом, оцениваемыми через анализ защищённости пользователей информационных систем при многоходовых социоинженерных атаках. Подход к оценке защищённости пользователей при многоходовых социоинженерных атаках, а также к расчёту оценок вероятности распространения атаки от пользователя к пользователю представлен в [4]. Отметим, что в [4] рассмотрены неориентированные графы, однако оценка вероятности распространения атаки от пользователя к пользователю в прямом и обратном направлении может быть разной. Также в [4] не рассматриваются вопросы поиска наиболее критичных траекторий распространения атаки. Как правило, существует несколько возможных траекторий развития атаки от одного пользователя к другому и вероятности успеха распространения атаки по каждой из них будут иметь разные значения. В связи с этим, актуальной видится проблема выявления наиболее критичных

[1] Стажёр-исследователь, студент.
[2] Научный сотрудник.
[3] Доктор физико-математических наук, главный научный сотрудник, профессор.

траекторий, вероятность прохождения по которым будет наиболее высокой. Одна из задач в контексте этой проблемы связана с выявлением наиболее критичных траекторий распространения атаки от одного пользователя к другому.

Математическая формализация исследуемой проблематики. Автоматизированный поиск критичных траекторий

Анализ возможных траекторий предлагается производить на ориентированном социальном графе сотрудников компании. Под социальным графом сотрудников компании будем понимать граф, вершины которого соответствуют сотрудникам компании, а рёбра — связям между сотрудниками. Формализуя описанное, пусть дан граф: $G=(U,E)$, где $U=\{User_i\}_{i=1}^{n}$ — множество вершин (пользователей), $E=\{u_i,u_j,p_{i,j}\}_{1\le i,j\le n,i\ne j}$ — множество упорядоченных троек с заданной оценкой вероятности распространения атаки от пользователя к пользователю — $p_{i,j}$. Заметим, что равенство $p_{i,j}$ и $p_{j,i}$ не предполагается. То есть вероятность распространения атаки от первого пользователя ко второму может отличаться от вероятности распространения атаки в другую сторону — от второго к первому.

Отметим, что оценка вероятности распространения социоинженерной атаки от пользователя к пользователю, согласно [4] рассчитывается следующим образом: $p_{i,j}=1-\prod_t\left(1-p_t^{i,j}\right)^{n_t}$, где $p_t^{i,j}$ — оценка вероятности успеха социоинженерной атаки злоумышленника на пользователя по t-ой связи, n_t — число эпизодов. В рассматриваемом частном случае модели, учитывающей сведения, извлекаемые из социальных сетей, $p_{i,j}>0$, рёбра, где оценка вероятности $p_{i,j}=0$ исключаются из итогового социального графа.

Задача поиска наиболее критичной траектории многоходовой социоинженерной атаки от $User_i$ до $User_j$ сводится к задаче нахождения в графе элементарного пути (простого и без циклов) между этими вершинами. Причём путь должен быть таким, что произведение оценок вероятностей переходов от пользователя к пользователю, входящих в него, максимально. Будем называть оценку вероятности успеха многоходовой социоинженерной атаки, которая представляет собой произведение оценок вероятностей распространения атаки от пользователя к пользователю и прямой атаки на первого пользователя, длиной пути.

Для упрощения, не умаляя общности, рассмотрим граф $G=(U,E')$, где $U=\{User_i\}_{i=1}^{n}$ — множество вершин (пользователей), $E'=\left\{u_i,u_j,\frac{1}{p_{i,j}}\right\}_{1\le i,j\le n,i\ne j}$ — множество упорядоченных троек, где каждой паре пользователей u_i,u_j сопоставлено число — $\frac{1}{p_{i,j}}$. Заметим, что в этом случае если $p_{i,j}\ge p_{l,k}$, то $\frac{1}{p_{i,j}}\le\frac{1}{p_{l,k}}$, а длина пути будет вычисляться следующим образом $\frac{1}{p_{ml}}=\frac{1}{p_m}\prod_{i=m}^{l-1}\frac{1}{p_{i,i+1}}$, где p_{ml} — оценка вероятности успеха прохождения атаки от пользователя m до пользователя l, p_m — оценка вероятности успеха прямой социоинженерной атаки злоумышленника на пользователя, $p_{i,i+1}$ — соответствующая оценка вероятности распространения атаки на пользователя через другого пользователя. Таким образом, от задачи поиска пути с максимальной длиной перейдём к задаче поиска пути с минимальной длиной.

Чтобы применить алгоритмы нахождения минимального пути необходимо произвести некоторые преобразования. Воспользуемся основным логарифмическим тождеством: $\frac{1}{p_{ij}}=e^{\log\frac{1}{p_{ij}}}$. Тогда длина пути преобразуется следующим образом:

$$\frac{1}{p_m}\prod_{i=m}^{l-1}\frac{1}{p_{i,i+1}} = e^{\log\frac{1}{p_m}}\prod_{i=m}^{l-1}e^{\log\frac{1}{p_{i,i+1}}} = \exp\left\{\log\frac{1}{p_m}+\sum_{i=m}^{l-1}\log\frac{1}{p_{i,i+1}}\right\}.$$

Поскольку оценка p_m для всех траекторий, начинающихся с пользователя m, будет одинакова, то наша задача сведётся к поиску пути в котором: $\sum_{i=m}^{l-1}\log\frac{1}{p_{i,i+1}}$ — минимальна среди всех возможных траекторий начинающихся с пользователя m и заканчивающихся пользователем l. Таким образом, задача представляет собой стандартный поиск наикратчайшего пути в ориентированном графе без рёбер отрицательного веса.

Пусть n — число вершин в социальном графе (число сотрудников), m — число дуг в социальном графе. Рассмотрим некоторые алгоритмы поиска кратчайшего пути:

- Алгоритм Дейкстры. [5] Вычислительная сложность алгоритма: $O(n^2)$. Данный алгоритм хорошо работает на графах с небольшим числом вершин. Социальные графы сотрудников компаний, как правило, достаточно большие.
- Алгоритм Беллмана-Форда. [6] Вычислительная сложность: $O(mn)$. Алгоритм для поиска кратчайшего пути от одной вершины графа до всех остальных, но при этом работает и с дугами отрицательного веса. В контексте рассматриваемой задачи условия алгоритма избыточны, что может привести к потере времени и лишним затратам памяти.
- Модификация алгоритма Дейкстры на основе фибоначчиевой кучи. [5] Сложность алгоритма: $O(n\log n+m)$. Однако в асимптотических оценках трудоемкости данной модификации скрыты константы, значение которых зачастую оказывается велико. Также, недостатком в нашем случае является поиск кратчайших путей от одной вершины графа до всех остальных.
- Алгоритмы Флойда-Уоршелла [5] имеет вычислительную сложность $o(n^3)$. Данный алгоритм хорошо применим на плотных графах. Однако является чересчур избыточным для целей исследуемой задачи, так как решает задачу о нахождении кратчайшего пути для всех пар вершин графа.
- Алгоритм A* [7] является эвристикой классического алгоритма Дейкстры. Чаще всего используется для поиска кратчайшего пути между точками в евклидовом пространстве. Для исследуемой задачи удобен тем, что позволяет осуществить поиск кратчайшего расстояния только между двумя вершинами, а не между всеми. Однако использование данного алгоритма в рамках решаемой задачи затруднительно, в связи с тем, что для него необходимо подобрать правильную эвристическую функцию $h(x)$. Данная функция должна не превосходить расстояния до целевой вершины и быть наиболее точной. Кроме того, недостатком A* является существенное потребление ресурсов памяти.

С учётом вышеперечисленного и в связи с тем, что задача не всегда предполагает работу с разреженным графом, была выбрана реализация алгоритма Дейкстры, но с дополнительными условиями. А именно, для уменьшения вычислительной сложности алгоритма введём пороговое значение. То есть если длина пути от начальной вершины до обрабатываемой в данный момент вершины становится меньше заданного порога, то исключим текущую вершину из рассмотрения. Кроме того, если путь до конечной вершины уже найден, то по условиям задачи дальнейших вычислений проводить не требуется, и алгоритм может быть завершён.

Заключение

Основным результатом работы является математически обоснованное сведение вычислительно затратной задачи к более простому случаю, который является уже широко исследованным. Что позволило, произвести более тщательное рассмотрение возможных реализаций и выбрать наиболее подходящую в рамках изучаемого вопроса. В дальнейшем планируется обобщение рассматриваемой задачи. А именно, поиск наиболее критических траекторий с учётом дополнительных условий, в частности, ресурсов злоумышленника. Также для моделирования и

оценки вероятностей успеха распространения многоходовых социоинженерных атак может быть использован аппарат байесовских сетей [8].

Работа выполнена в рамках проекта по государственному заданию СПИИРАН № 0073-2014-0002, при финансовой поддержке РФФИ, проект №18-01-00626 – Методы представления, синтеза оценок истинности и машинного обучения в алгебраических байесовских сетях и родственных моделях знаний с неопределенностью: логико-вероятностный подход и системы графов; проект № 18-37-00323 – Социоинженерные атаки в корпоративных информационных системах: подходы, методы и алгоритмы выявления наиболее вероятных траекторий; проект №18-37-00340 – Методы анализа устойчивости структуры социальных связей пользователей информационной системы к социоинженерным атакующим воздействиям злоумышленника на основе применения генетических алгоритмов.

ЛИТЕРАТУРА

1. ЦБ ожидает роста активности мошенников, использующих социальную инженерию. https://ria.ru/economy/20171213/1510861611.html [Дата обращения: 07.05.2018].
2. По следам CyberCrimeCon 2017: Тенденции и развитие высокотехнологичной преступности. https://habr.com/company/group-ib/blog/341812/ [Дата обращения: 16.04.2018].
3. Азаров А.А., Тулупьева Т.В., Суворова А.В., Тулупьев А.Л., Абрамов М.В., Юсупов Р.М. Социоинженерные атаки: проблемы анализа / Под общ. ред. Р. М. Юсупова. — СПб.: Наука, 2016. — 349 с.
4. Абрамов М.В., Тулупьев А.Л., Сулейманов А.А. Задачи анализа защищенности пользователей от социоинженерных атак: построение социального графа по сведениям из социальных сетей // Научно-технический вестник информационных технологий, механики и оптики. 2018. Т. 18. № 2. С. 313–321. doi: 10.17586/2226-1494-2018-18-2-313-321.
5. Кормен Т., Лейзер Ч. Алгоритмы. Построение и анализ — М.: МЦНМО, 2002. — 960 с.
6. Левитин А. Алгоритмы: введение в разработку и анализ — М.: Издательский дом "Вильямс 2006. - 576 с.
7. Рассел С., Норвиг П. Искусственный интеллект. Современный подход. 2-е изд. — М.: Вильямс, 2007. — 1410 с. — ISBN 5-8459-0887-2, 0-13-790395-2, 978-5-8459-0887-2.
8. Харитонов Н.А., Березин А.И. Синтез математического представления ациклической алгебраической байесовской сети // Сборник докладов Международной конференции по мягким вычислениям и измерениям (SCM-2018). Санкт-Петербург. Том 1-2. Т. 1. 2018. С. 141–143.

A.O. Khlobystova, M.V. Abramov, A.L.Tulupyev (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science, Saint-Petersburg State University, St. Petersburg)

Identification of the most probable trajectors of socio-engineering attacks in management of risks associated with users

The purpose of this work is to develop approaches to solving the problem of identifying the most critical trajectories of the propagation of multi-path socio-engineering attacks and their subsequent implementation. Formally, the problem reduces to finding a path in the graph, in which the product of the weights of all edges entering the given path is maximal. The difficulties that arise in more detailed consideration of the problem are that the graph is a complete graph with a large number of vertices, which makes the process resource-consuming, and it is also impossible to apply known algorithms to find the paths in the graph because the edges weights are probability. The last mentioned problem can be solved by considering the values of the reverse edges to the values of the weights, which makes it possible to go to the search for the minimum path in the graph, it is also possible to logarithmize the weights, so that their sum will be considered instead of the product of the weights. Thus, the problem reduces to a standard search for the shortest path in a graph. The density of the graph is proposed to be reduced by eliminating arcs with small probabilities from consideration. The main result of the work is finding the optimal algorithm for searching for the most vulnerable trajectories and its software implementation.

АВТОМАТИЗАЦИЯ ОЦЕНКИ СТЕПЕНИ ВЫРАЖЕННОСТИ ПСИХОЛОГИЧЕСКИХ ОСОБЕННОСТЕЙ ПОЛЬЗОВАТЕЛЕЙ В УПРАВЛЕНИИ КАДРАМИ: ТЕСТИРОВАНИЕ В СОЦИАЛЬНОЙ СЕТИ

Р. А. Шаламов[1], **М. В. Абрамов**[2], **Т. В. Тулупьева**[3]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук,
199178, С.-Петербург, 14-я лин. В.О., 39;
Санкт-Петербургский государственный университет,
198504, Санкт-Петербург, Петергоф, Университетский пр., 28,
shalamov.ra@gmail.com, mva16@list.ru, tvt100a@mail.ru*)

Аннотация

Для защиты критичной информации от программно-технических атак в настоящее время разработано множество методов и программных комплексов, сильно затрудняющих, а порой и делающих невозможным получение доступа к ней злоумышленником. Однако, доступ к системе можно получить не только путем ее программно-технического взлома, но и с помощью воздействия на пользователя данной системы, имеющего доступ к критичной информации. Данная работа является постановочно, её цель заключается в описании архитектуры приложения, предназначенного для сбора информации о пользователе на основе текстовой информации с его личной страницы и результатах прохождения им определенных психологических тестов (представленных в приложении), с помощью которых в дальнейшем можно будет диагностировать возможные уязвимости пользователя информационной системы и проводить профилактику социоинженерных атак на данного пользователя.

Ключевые слова: социоинженерные атаки, защита пользователя, критичная информация, анализ защищённости, социальные сети, критичные документы, анализ социальных сетей, информационная безопасность.

Введение

В настоящее время информационные технологии получили широкое распространение во всех сферах жизни человека. Согласно статистике, активно пользуются интернетом 3.01 млрд. человек, социальными сетями — 2.078 млрд. человек [1]. Такое масштабное использование информационных систем привело к оперированию большими потоками данных, которые необходимо безопасно хранить, обрабатывать, анализировать и т.д. Большая часть информации сегодня хранится в электронном виде.

Информационные системы используются как в личных, так и в рабочих целях (например, как часть коммерческой организации). Информационная система организации может содержать критичную информацию, о безопасности которой необходимо заботиться.

Для защиты критичной информации в настоящее время разработано множество методов и программных комплексов, сильно затрудняющих, а порой и делающих невозможным получение доступа к ней злоумышленником, производящим программно-техническую атаку. Однако, доступ к системе можно получить не только путем ее взлома, но и с помощью воздействия на пользователя данной системы, имеющего доступ к критичной информации. В настоящее время известно, что одним из наиболее уязвимых мест большинства информационных систем является ее пользователь [2, 3], поэтому воздействие на него с целью, например, получения несанкционированного доступа к системе является наиболее приоритетной задачей хакеров. Такое воздействие на пользователя с целью проникновения в информационную систему называется социоинженерной атакой.

На данный момент уже предложен подход к анализу защищенности пользователя от социоинженерной атаки [4], основанный на нечетком и вероятностном подходах. Однако, для более подробной диагностики уязвимости пользователя необходимо большее количество данных, анализируя которые, можно будет предложить более качественные оценки в профиле уязвимости пользователя.

[1] Стажёр-исследователь, студент.
[2] Научный сотрудник, старший преподаватель.
[3] Кандидат психологических наук, доцент, старший научный сотрудник, доцент.

Описание архитектуры приложения. Обзор стека используемых технологий.

Разрабатываемое приложение типа IFrame [5], т.е. представляет собой элемент сайта, загружаемый через страницу «ВКонтакте». На основной странице приложения предложено выбрать один из нескольких психологических тестов [6], который впоследствии необходимо пройти (рис. 1).

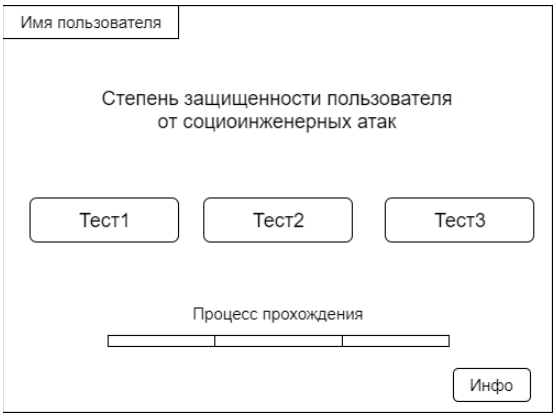


Рис. 1. Схема интерфейса приложения

Пользователю, устанавливающему приложение, необходимо предоставить доступ к основной информации со своей личной страницы «ВКонтакте», которая в дальнейшем будет записана в базу данных вместе с результатами прохождения тестов.

Ниже представлена диаграмма компонент разрабатываемого приложения (рис.2).

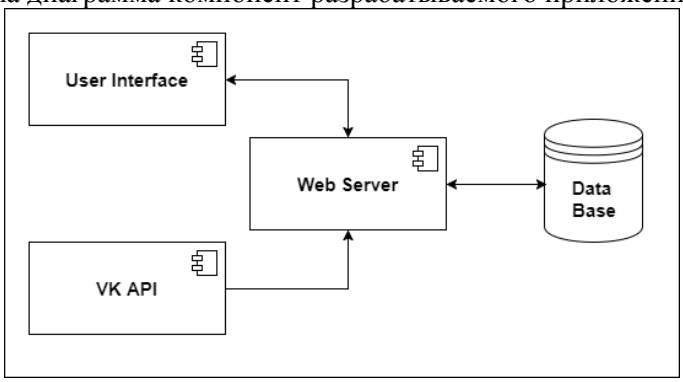


Рис. 2. Диаграмма компонент приложения

Пользовательский интерфейс (User Interface) — интерфейс, обеспечивающий обмен информации между человеком и информационной системой. В контексте данной работы пользователь взаимодействует со страницей «ВКонтакте» и проходит отображаемые на ней психологические тесты. Результаты тестов и данные пользователя отправляются веб-серверу для дальнейшей обработки и выдачи пользователю результата о прохождении.

Серверная составляющая (Web Server) подразумевает «прослойку» между пользовательским интерфейсом и существующей базой данных. Веб-сервер будет содержать функции, обрабатывающие данные о пользователе и пройденных им тестах для удобного представления их в базе данных.

Для хранения информации, которая будет анализироваться в дальнейшем, необходимо создание базы данных (Data Base). С помощью неё будет анализироваться вся полученная о пользователе информация, результаты пройденных тестов, а также сами тесты, которые будут передаваться в пользовательский интерфейс.

Для сбора информации о пользователе используется VK API - набор готовых функций, с помощью которых можно взаимодействовать с социальной сетью «ВКонтакте».

Как было описано ранее, приложение будет представлять собой веб-страницу, размещенную в социальной сети «ВКонтакте», поэтому технологии для его разработки можно разбить на три части: пользовательский интерфейс, серверная составляющая и база данных.

Для разработки пользовательского интерфейса будет использован язык разметки документов HTML (создание каркаса веб-сайта), язык описания внешнего вида документа CSS (стилизация веб-сайта), язык программирования JavaScript (для придания интерактивности веб-сайту).

Для разработки серверной части был выбран фреймворк ASP.NET MVC[7]. Основными преимуществами данного инструмента перед другими являются:

- Использование архитектурного паттерна MVC.
- Разнообразие библиотек.

Для создания и управления базой данных была выбрана система управления реляционными базами данных Microsoft SQL Server. Среди преимуществ данной технологии можно выделить быстродействие и простоту при работе с небольшими проектами, а также удобную интеграцию с технологией ASP.Net MVC.

Заключение

Из анализа тематической литературы было выявлено, что для оценки вероятности успеха социоинженерной атаки необходимо агрегировать большое количество данных, в том числе связанных с психологическими и иными особенностями личности. Для построения профиля уязвимостей необходимо выявлять психологические особенности пользователя, источником информации для которых могут служить социальные сети. Для этого решается задача автоматизированного сбора информации о пользователя с целью дальнейшего анализа и выявления психологических особенностей. В данной работе представлены результаты первого этапа этой работы, связанные с разработкой архитектуры приложения для социальной сети «ВКонтакте».

В качестве дальнейшего развития выбраны следующие направления.

- Реализация приложения на основе разработанной архитектуры.
- Публикация приложения в социальной сети «ВКонтакте».
- Сбор данных о пользователях с помощью данного приложения для дальнейшего анализа.

Работа выполнена в рамках проекта по государственному заданию СПИИРАН № 0073-2014-0002, при финансовой поддержке РФФИ, проект №18-01-00626 – Методы представления, синтеза оценок истинности и машинного обучения в алгебраических байесовских сетях и родственных моделях знаний с неопределенностью: логико-вероятностный подход и системы графов; проект № 18-37-00323 – Социоинженерные атаки в корпоративных информационных системах: подходы, методы и алгоритмы выявления наиболее вероятных траекторий. Если работа проводилась при поддержке какой-либо организации, то в конце текста доклада необходимо сделать ссылку на эту организацию; проект №18-37-00340 – Методы анализа устойчивости структуры социальных связей пользователей информационной системы к социоинженерным атакующим воздействиям злоумышленника на основе применения генетических алгоритмов.

ЛИТЕРАТУРА

1. Digital, Social & Mobile in 2015 [Электронный ресурс]. — URL: https://wearesocial.com/sg/special-reports/digital-social-mobile-2015 [Дата обращения: 22.04.2018].
2. Котенко И.В., Юсупов Р.М. Перспективные направления исследований в области компьютерной безопасности. Защита информации. — Инсайд. 2006. № 2. — С. 46.
3. Азаров А.А., Тулупьева Т.В., Суворова А.В., Тулупьев А.Л., Абрамов М.В., Юсупов Р.М. Социоинженерные атаки: проблемы анализа. — СПб.: Наука, 2016. — 349 с.
4. Азаров А.А., Тулупьева Т.В., Тулупьев А.Л. Прототип комплекса программ для анализа защищенности персонала информационных систем, построенный на основе фрагмента профиля уязвимостей пользователя. — Труды СПИИРАН, Вып. 21., 2012. — с. 21- 40.
5. IFrame-приложения [Электронный ресурс]. — URL: https://vk.com/dev/IFrame_apps [Дата обращения: 15.05.2018].
6. Капустина А. Н. Многофакторная личностная методика Р. Кэттелла. - СПб.: Речь, 2001.
7. ASP.NET MVC | The ASP.NET Site [Электронный ресурс]. — URL: https://www.asp.net/mvc [Дата обращения: 30.05.2018].

R.A. Shalamov, M.V. Abramov, T.V. Tulupyeva (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science, Saint-Petersburg State University, St. Petersburg)

Automation of evaluation degree of expression of user's psychological features in management of human resources: testing in the social network.

To protect critical information from software and hardware attacks, there are now many methods and software packages that make it difficult, and sometimes impossible, to gain access to it by an attacker. However, access to the system can be obtained not only by hacking it, but also by influencing the user of this system, which has access to critical information. The purpose of this paper is to describe the architecture of an application designed to collect information about a user based on textual information from his personal page and the results of passing certain psychological tests (presented in the application) with which to further diagnose possible vulnerabilities of the user of the information system and conduct prevention of socio-engineering attacks on this user.

Секция 7
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ЧЕЛОВЕКО-МАШИННОГО ВЗАИМОДЕЙСТВИЯ

СТАТИСТИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ ФОРМАНТ РАЗЛИЧНЫХ ДИКТОРОВ ПРИ ПРОВЕДЕНИИ ФОНОСКОПИЧЕСКИХ ЭКСПЕРТИЗ МЕТОДОМ ФОРМАНТНОГО ВЫРАВНИВАНИЯ

И. А. Гураков[1], **Е. Ю. Костюченко**[2], **Д. И. Новохрестова**[3], **А. А. Шелупанов**[4]
(*ФГБОУ ВО «Томский государственный университет систем управления и радиоэлектроники», Томск, 634050, пр. Ленина, д. 40, key@keva.tusur.ru*)

Аннотация

Цель данной работы - получение статистического распределения формантных частот для нескольких дикторов и проведение анализа на основе полученных распределений о допустимости сделанных ранее предположений и необходимости их корректировки. Показано что для одной фонемы и одного диктора предположение о равномерности распределения формантных частот не подтверждается и необходима дальнейшая проверка такого предположения для множества дикторов. Определены частоты для выравненных и полностью совпавших участков для 6 различных гласных звуков в рамках одного диктора и двух различных дикторов. Показана как высокая вероятность несовпадения для выравненного участка одного диктора, так и возможность полного совпадения формант для разных дикторов на полностью фонетически идентичных фрагментах. Данные результаты могут быть использованы при построении методики фоноскопического исследования на основе статистических распределений формантных частот.

Ключевые слова: форманта, фоноскопическая экспертиза, распределение частот, формантное выравнивание.

Введение

Одним из методов проведения фоноскопических экспертиз является метод формантного выравнивания. В рамках этого метода экспертом сперва находятся участки, содержащие одинаковые с фонетической точки зрения фрагменты, далее проводится выделение резонансных частот речевого тракта (формант) для этих фрагментов и выбираются только те из них, для которых частоты первой и второй формант совпадают (выравненные фрагменты). На следующем для выравненных фрагментов проводится сравнение третьей и четвертой формант и если и они совпадают – то такой фрагмент является аргументом в пользу утверждения что на сравниваемые фрагменты были произнесены одним и тем же диктором. Однако прямое применение такого подхода порождает ряд вопросов: какие форманты считать совпадающими (какой может быть максимально допустимый разброс), сколько необходимо совпавших фрагментов и какое влияние оказывают фрагменты с несовпадающими частотами формант, имеется ли их максимальная предельно допустимая доля для принятия решения о несовпадении дикторов и чему она равна. В части этих утверждений классический метод формантного выравнивания основывается на предположении о равномерном распределении частот формант, однако в реальной жизни такое предположение является как минимум спорным.

Целью данной работы является получение статистических распределения формантных частот для нескольких дикторов и проведение анализа на основе полученных распределений о допустимости сделанных ранее предположений и необходимости их корректировки. В рамках сделанного анализа показано что для одной фонемы и одного диктора предположение о равномерности распределения формантных частот не подтверждается и необходима дальнейшая проверка такого предположения для множества дикторов. Определены частоты для выравненных и полностью совпавших участков для 6 различных гласных звуков в рамках одного дикто-

[1] Инженер.
[2] Кандидат технических наук, младший научный сотрудник.
[3] Техник.
[4] Доктор технических наук, ректор.

ра и двух различных дикторов. Показана как высокая вероятность несовпадения для выравненного участка одного диктора, так и возможность полного совпадения формант для разных дикторов на полностью фонетически идентичных фрагментах. Данные результаты могут быть использованы при построении методики фоноскопического исследования на основе статистических распределений формантных частот.

Построение статистического распределения формантных частот

Постановка задачи. В криминалистике часто возникает потребность установления личности человека по имеющейся звукозаписи. С этой целью было разработано несколько методик идентификации лиц на основании физических параметров звучащей речи с применением ЭВМ. Однако большинство этих методик требует наличия специализированного оборудования, позволяющего непосредственно обрабатывать звуковые сигналы [1]. При этом, исследование звукозаписи перцептивными методами все равно является обязательным.

При проведении экспертиз имеется возможность отталкиваться от множества спектральных параметров, работать с усредненным спектром мощности, спектрами более высоких порядков, использовать частоты основного тона [2-6]. Одним из важнейших параметров являются значения формантных частот. Форманта – собственная частота резонатора голосового аппарата. После того, как звук зародился в голосовых связках, он проходит через сеть резонаторов [7]. Резонаторы — в основном, это полости, которые в результате физического явления совпадения частот колебаний (физическое явление резонанса) увеличивают силу первичного звука, рождаемого на связках, и окрашивают его [8]. Собственная частота резонатора зависит от его положения.

Однако, при исследовании речи, нет возможности напрямую измерить параметры голосового аппарата. Поэтому при анализе голоса [9-11], под формантами [12] принято понимать выраженные пики мощности спектра (Рис. 1).

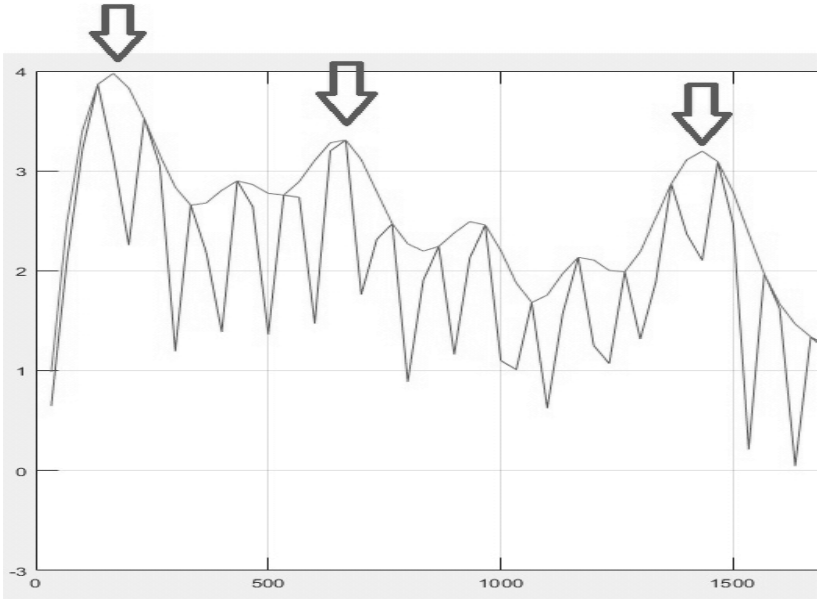


Рис. 1. Выраженные пики мощности спектра, зависимость интенсивности от частоты, Гц

На базе формант может проводится сравнение формантных спектров подобных артикуляторных событий, а также анализ особенностей формантных структур [13]:

В данной работе проводится исследование метода формантного выравнивания как одного из широко используемых на практике в рамках проведения криминалистических экспертиз [14]. По сути методика основывается на поиске фонетически идентичных фрагментов с совпадающими частотами первой и второй формант и определении совпадения/несовпадения частот третьей и четвертой формант на таких участках. Методика имеет несколько моментов, требующих рассмотрения:

- согласно стандартной методике, специалист должен проанализировать аудиозапись, найти подобный артикуляторные события, найти в них закономерность и выбрать промежуток 10-

30мс, на основе которого будут выделяться форманты. Однако нет никакой гарантии, что эти, взятые «на глаз» промежутки действительно будут в достаточной степени совпадать, а, даже если и будут, то даже у одного диктора в одинаковых слогах формантные траектории могут сильно различаться раз от раза;

- интервал, в пределах которого можно говорить о совпадении или несовпадении формант. Полное совпадение четырех формант – довольно редкое совпадение даже для одного диктора, с другой стороны, при желании можно найти совпадения и у разных дикторов;

- традиционная методика утверждает, что достаточно найти 7-10 выравненных участков, чтобы утверждать о совпадении дикторов, однако никак не освещается тот факт, что выравненных участков с несовпадающим полным набором формант может быть гораздо больше.

Задача настоящей работы состоит в разработке автоматизированного алгоритма измерения частот формант (или кандидатов в форманты), построения их распределения, проведения анализа этих распределений, выявления закономерностей и наличия/отсутствия аргументов в пользу модификации имеющегося метода.

Выделение формантных частот. Первой проблемой, с которой приходится столкнутся при применении метода формантного выравнивания – это выделение частот формант и построение динамики их изменения (так называемых формантных траекторий). В рамках данной работы для выделения формантных частот и кандидатов в формантные частоты (с учетом потенциальных ошибок готовится избыточный список, учитывающий возможность того, что ложная, или как ее еще называют, «фантомная» форманта при регулярном возникновении тоже может нести информацию о дикторе). В рамках работы данного алгоритма выделения частот по сути на выходе получается последовательность частот, на которых наблюдаются максимумы. Однако, в отличии от традиционной методики, для дальнейшей обработки используется не значение одной частоты максимума и не их последовательность, а их статистическое распределение. Это позволяет объединять фонетически однородные участки в одну выборку, основанную на большем количестве отдельных значений, и, как следствие, обеспечивающую большую значимость получаемых на ее основе значений как по сравнению с отдельными меньшими выборками, так и с отдельными значениями, получаемыми зачастую не только с использованием объективных оценочных методов.

Укрупненно алгоритм выделения частот формант может быть представлен в виде следующей последовательности шагов:

1. Задать пустую матрицу найденных формант.
2. Задать диапазоны поиска для каждой из 4 формант.
3. Задать пороговое значение разниц между максимумами *ThI*
4. Построить огибающую спектра.
5. Цикл по 4 формантам i=1..4
   - 5.1 Найти 3 положения максимумов интенсивности огибающей в i-м диапазоне *I1*, *F1*, *I2*, *F2*, *I3*, *F3*
   - 5.2 Выбор из 9 возможных ситуаций
     - 5.2.1 Если Abs(*I2-I1*)<*ThI* и Abs(*I2-I3*)<*ThI*, то i-я форманта не определилась, добавить в матрицу формант *Fi*=0, переход на 5.
     - 5.2.2 Если Abs(*I2-I1*)<*ThI* и (*I3-I2*)>=*ThI*, то добавить в матрицу формант *F3* как i-ю форманту, переход на 5.
     - 5.2.3 Если Abs(*I2-I1*)<*ThI* и (*I2-I3*)>=*ThI*, то продублировать все строки матрицы, в первую половину строк добавить *F1*, во вторую половину строк добавить *F2* как i-ю форманту, переход на 5.
     - 5.2.4 Если (*I2-I1*)>=*ThI* и (*I3-I2*)>=*ThI*, то добавить в матрицу формант *F3* как i-ю форманту, переход на 5.
     - 5.2.5 Если (*I2-I1*)>=*ThI* и Abs(*I3-I2*)<*ThI*, то продублировать все строки матрицы, в первую половину строк добавить *F2*, во вторую половину строк добавить *F3* как i-ю форманту, переход на 5.
     - 5.2.6 Если (*I2-I1*)>=*ThI* и (*I2-I3*)>=*ThI*, то добавить в матрицу формант *F2* как i-ю форманту, переход на 5.

5.2.7 Если (*I1-I2*)>=*ThI* и (*I3-I2*)>=*ThI*, то продублировать все строки матрицы, в первую половину строк добавить *F1*, во вторую половину строк добавить *F3* как *i*-ю форманту, переход на 5.

5.2.8 Если (*I1-I2*)>=*ThI* и Abs(*I3-I2*)<*ThI*, то добавить в матрицу формант *F1* как *i*-ю форманту, переход на 5.

5.2.9 Если (*I1-I2*)>=*ThI* и (*I2-I3*)>=*ThI*, то добавить в матрицу формант *F1* как *i*-ю форманту, переход на 5.

6. Вывод матрицы формант.

Графически шаг выбора форманты 5.2 может быть проиллюстрирован рисунком 2, где для каждой ситуации выделена выбираемая форманта.

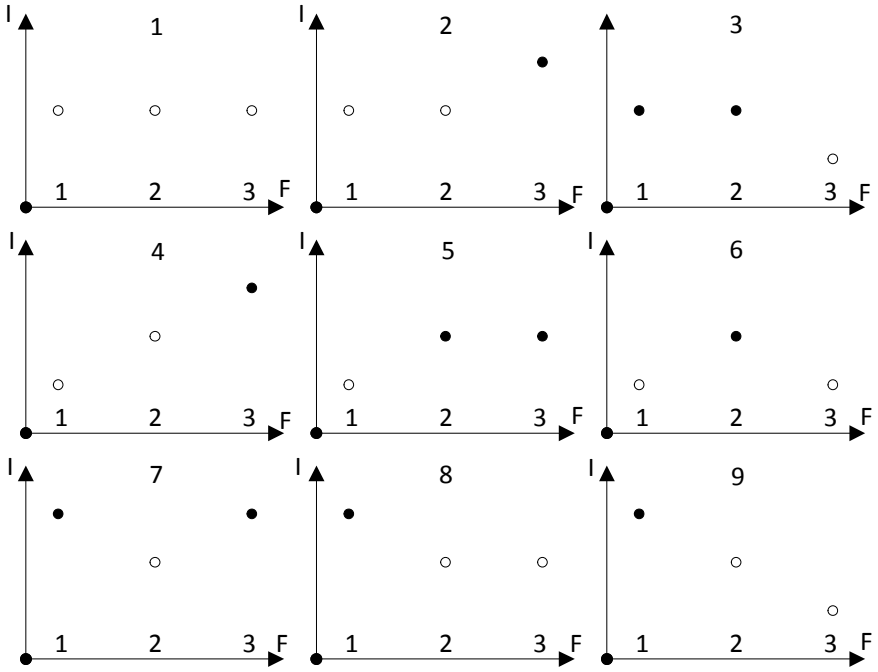


Рис. 2. Выбор формант из потенциальных кандидатов. Черные точки – выбранные форманты в зависимости от номера ситуации (условия)

Для построения огибающей спектра (шаг 4) было использовано преобразование Гилберта [15].

В результате работы алгоритма получается матрица комбинаций значений четырех первых формант, при этом число таких комбинаций может превосходить одну из-за ситуаций 3,5 и 7, при которой кандидатами на роль форманты считаются оба возможных значения частоты, а пропуски потенциальных формант появляются только в ситуации примерного равенства трех максимальных значений амплитуд пиков в рамках рассматриваемого диапазона. Степень равенства задается размером порогового значения *ThI*. На основании полученных данных могут быть построены статистические распределения формант и кандидатов на их роль для последующего принятия решения об идентичности личности говорящего на основании близости статистических распределений.

Поиск выравненных участков. Для проведения эксперимента был разработан программный модуль для обработки массива формант, сформированных на выходе вышеописанного алгоритма. Задачей данного модуля был поиск участков с совпадающими частотами формант (с заданной точностью). Участки с совпадающей первой парой формант по сути являются пригодными для использования в рамках метода формантного выравнивания (конечно, при условии, что это совпадение достигается не в одной случайной точке, а на некотором интервале). Однако даже с учетом этого ограничения подобный модуль существенно повышает автоматизацию поиска таких участков за счет отбрасывания фрагментов с явным расхождением частот формант. Кроме того, данный программный модуль может быть использован и при поиске совпадений в рамках формант высших порядков. Интерфейс работы данного модуля при проведении сравнения представлен на рисунке 3. Вся форма разбита на 2 половины, где с левого края находится информация по первому слогу, с правого симметрично относительно центра информация по второму слогу. Области дан-

ных по порядку: задание областей поиска формант, найденные в рамках слога форманты. По середине сверху – фрагменты с совпавшей первой парой формант, снизу – фрагменты с совпавшими всеми четырьмя формантами. 4 поля снизу – допуски на совпадение каждой из формант.

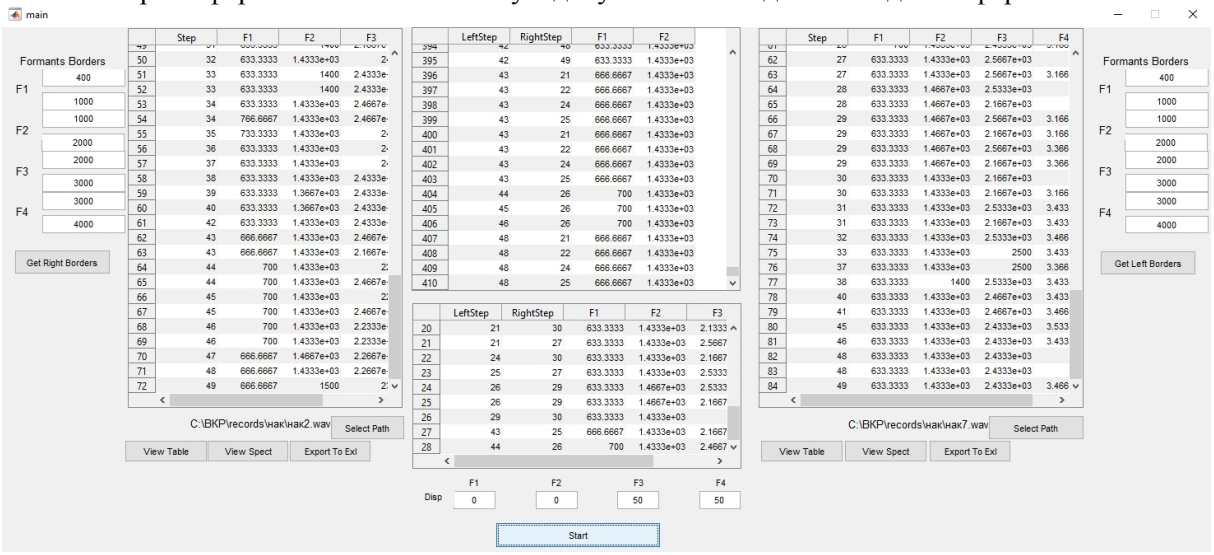


Рис. 3. Интерфейс модуля выделения формант двух звукозаписей и поиска участков с совпадающими двумя и четырьмя формантами

Дальнейшая обработка проводится по сути на основе данных об участках с совпадающими частотами формант.

Данные для проведения эксперимента. Для проведения исследования предложенного методы была сформирована предварительная база записей со следующими характеристиками.

1. Материал записывался в слоговой форме. Логично, что при проведении реального исследования такая форма будет непригодна, однако для предварительной проверки она является наиболее подходящей. Если не будет предпосылок к применимости метода на отдельных слогах, то говорить о его более общей применимости не приходится. Как следствие, формантная структура строилась на основе ударных гласных (других в рамках слога не может быть в принципе), находящихся в полностью идентичной позиции.
2. Начальное количество дикторов было равно четырем, однако в качестве результатов для демонстрации были выбраны два диктора с наиболее пересекающимися частотами формант, которые по сути являются наиболее худшим случаем из рассматриваемых.
3. Объем записей: 6 слогов «нак», «нок», «нук», «нык», «нэк» и «ник». По 20 записей от каждого из дикторов на каждый слог. Общий объем предварительной базы данных составил 480 записей.

Анализ полученных результатов. По итогам эксперимента были получены следующие основные результаты:

1. При проведении исследования методом формантного выравнивания при определении статистических оценок точности метода одно из предположений, на котором по сути базируются таковые оценки, заключатся в равномерности распределения частот каждой из формант. В частности, это следует из следующего фрагмента с описанием метода: «Совпадение второй, отличной по частному положению четверки формант говорит о том, что у этих двух дикторов вокальные тракты совпали не только при одной конфигурации, но и при другой, отличной от первой. Такой факт говорит о том, что вокальные тракты данных дикторов довольно близки по размерам и пропорциям. Если остаться на позиции, что возможно различение не более 5 градаций четвертой форманты, то вероятность случайной реализации такого события 1/25. [14]». Видно, что вероятности градаций формант считаются одинаковыми. При этом, нет точной информации о конкретных диапазонах этих градаций в методике нет.

   В рамках данной работы были построены распределения для первых четырех формант. В частности, на рисунке 4 приведены гистограммы формантных частот нескольких пользователей для третьей (слева) и четвертой (справа) формант.

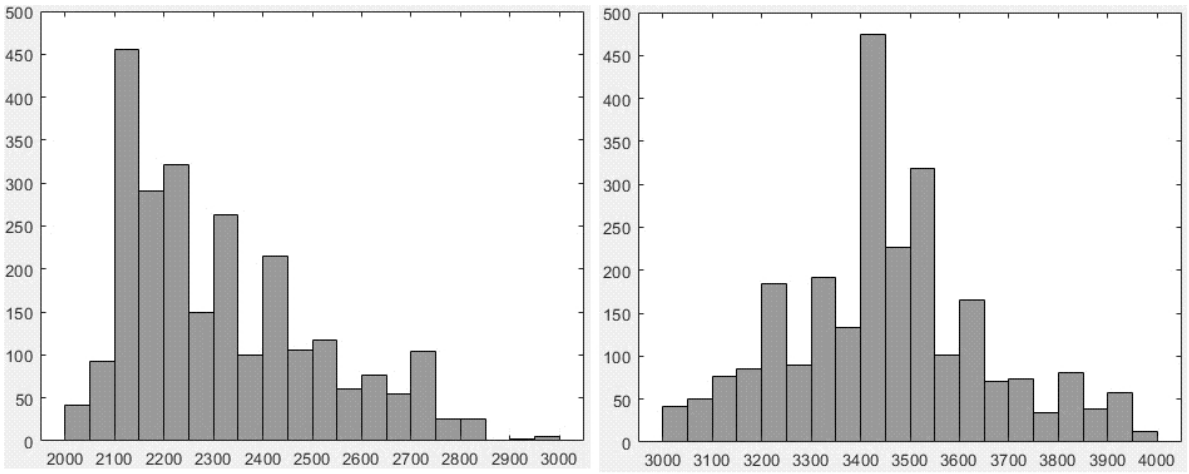


Рис. 4. Гистограммы частот встречаемости значений третьей (слева) и четвертой (справа) формант. Ось абсцисс - частота в Гц, ординат – частость, шт.

Видно, что распределение явно носит не равномерный характер. Кроме того, для возникновения в природе равномерного распределения нужны очень специфические условия (физические ограничения величины слева и справа при равновероятности значений между ними). Полученные результаты вместе с теоретическими предпосылками говорят в пользу того, что учет особенностей распределения повысит обоснованность методики.

2. Гистограммы встречаемости для разных пользователей носят явно разный характер. Так, информация на рисунке 4 иллюстрирует значения частот для третьей форманты для дикторов 1 (слева) и 2 (справа).

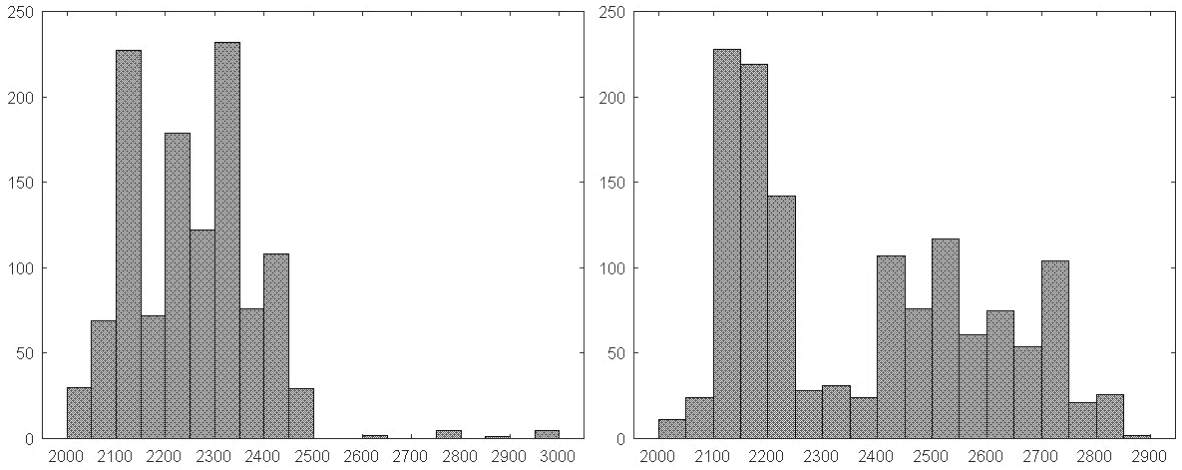


Рис. 5. Гистограммы частот встречаемости значений третьей форманты для первого (слева) и второго (справа) пользователей. Ось абсцисс - частота в Гц, ординат – частость, шт.

Видно, что зависимости носят качественно разный характер, что говорит в пользу необходимости учета не только отдельных значений, но и их распределений. Сравнения для остальных пар пользователей и четвертой форманты носит аналогичный характер. Понятно, что для получения количественных оценок вероятностей совпадения необходим значительно более полный эксперимент, явно можно утверждать о наличии полезной информации в полном распределении.

3. Значения формант могут иметь достаточно широкий разброс даже в рамках практически идентичных артикуляторных событий. Так, на рисунке 6 приведена гистограмма для полученных частот первой форманты для одного пользователя на базе 20 реализаций слога «нак». Даже если допусти что первый пик значений обеспечивается неотсечением «фантомных» формант, то второй пик, несмотря на кажущуюся простоту и стандартность ситуации, явно носит размытый характер и говорить об одном конкретном значении или даже о небольшом диапазоне значений не приходится. Явно видна необходимость учета особенностей именно распределения в целом.

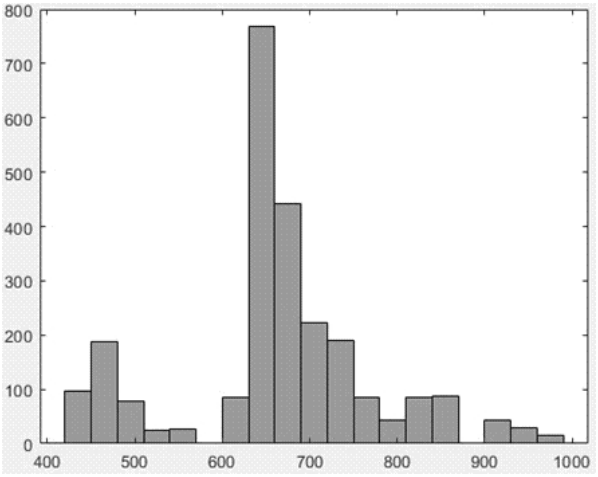


Рис. 6. Гистограмма частот встречаемости значений первой форманты для одного диктора, 20 реализаций слога «нак». Ось абсцисс - частота в Гц, ординат – частость, шт.

4. Рассмотрим более детально частоты совпадения для двух (выравненный участок) и четырех (совпавший участок). Результаты приведены в таблице 1. В рамках каждого слога в первом столбце значения для слогов первого диктора, во втором столбце – для первого слога первого диктора и второго слога второго диктора, в третьем столбце – для слогов второго диктора. Рассмотрим их более подробно.

Таблица 1

Результаты сравнения пары дикторов по всем гласным

	нак			нок			нук		
Среднее количествово выравненных участков	224,8	25,7	45,3	77,1	3,2	37,6	31,8	23,9	44,1
Среднее количествово комбинаций совпавших формант	16,7	0,7	3,6	7,6	0,02	6,3	3,4	0,5	6,9
Макс. количествово выравненных формант	866	264	254	344	36	244	122	119	112
Макс. количествово совпавших формант	93	6	24	147	6	32	13	6	26
Количествово пар записей без выравненных участков	6	181	4	0	257	42	6	15	0
Количество пар записей без совпавших участков	42	339	206	76	399	118	174	366	54
	нэк			ник			нык		
Среднее количествово выравненных участков	80,0	7,5	25,0	159,4	14,5	52,5	46,9	4,1	25,1
Среднее количествово комбинаций совпавших формант	9,1	0,6	4,2	6,5	0,8	5,5	4,1	0,3	4,7
Макс. количествово выравненных формант	352	80	205	457	164	230	260	68	224
Макс. количествово совпавших формант	41	9	36	22	6	32	28	7	21
Количествово пар записей без выравненных участков	2	189	60	6	151	16	14	228	36
Количество пар записей без совпавших участков	54	361	182	68	344	134	148	381	126

Измерялось количество участков с совпавшей первой парой формант (выравненные участки) и со всеми совпавшими четырьмя формантами (совпавшие участки). При сравнении указывались следующие разбросы формант: первая и вторая форманта – 0 Гц, третья и четвертая форманта – 33,3(3) Гц. Такой выбор обоснован тем, что первые две форманты менее подвержены изменениям при повторном произнесении слога одним диктором. Вторая и третья форманты напротив имеют свойство сильно искажаться, поэтому был взять минимально возможный неотрицательный разброс.

Видно, что даже в рамках одного пользователя количество совпавших участков значительно меньше количества выравненных. Это говорит о том, что даже на практически идентичном с фонетической точки зрения метериале для одного и того же диктора совпадение первой и второй формант отнюдь не влечет автоматическое совпадение третьей и четвертой формант.

Кроме того, видно, что задавшись целью можно найти полностью совпавшие участки даже для различных дикторов (пусть и в значительно меньшем количестве). Да, конечно, при ручном анализе такие участки будут практически наверняка будут отброшены на основе динамики изменения формант, однако оба этих факта говорят о необходимости учета именно множества значений, что возможно, в частности, за счет анализа полной гистограммы частот для каждой из фонем в идентичной позиции.

Если коротко подводить результаты преведенного эксперимента, то можно сделать основной вывод, что даже из таклой небольшой выборки сигналов ясно, что в статистических распределениях формантных частот содержится дополнительная информация, которая может повысить обоснованность получаемых выводов. Использование отдельных значений может приводить как к несовпадению верхних формант на выравненных участках одного пользователя, так и к полному несовпадению формантных частот разных пользователей.

Заключение

В рамках работы показаны недостатки существующего метода формантного выравнивания, заключающиеся в спорном предположении о равномерности распределения формантных частот внутри диапазона возможных значений. Показана необходимость учета не только отдельных значений формантных частот, но и их статистического распределения, полученного на основе анализа нескольких записей, содержащих близкие с фонетической точки зрения гласные. На данном этапе работа проведена на основе измерения формантных частот гласных фонем внутри изолированного слога. На следующем этапе предполагается провести аналогичное исследование и построить распределения для изолированных слов и для одинаковых реализаций слов, вырезанных из слитной речи. Кроме того, ведутся работы по увеличению количества дикторов для проверки утверждения о равномерности частот формант для множества дикторов.

Работа выполнена в рамках базовой части государственного задания министерства образования и науки Российской Федерации на 2017-2019 гг. Номер 8.9628.2017/8.9

ЛИТЕРАТУРА

1. Экспертиза видео- и звукозаписей [Электронный ресурс]. – Режим доступа: http://www.sudexpert.ru/possib/video.php (дата обращения 25.06.2018).
2. Wu C., Huang C., Chen H. Text-independent speech emotion recognition using frequency adaptive features. *Multimedia Tools and Applications*. 2018. P. 1-11.
3. Tabassum M., Aziz Jahan M.A., Rahman M.M., Mohamed S.B., Rashid M.A. Speaker independent speech recognition of isolated words in room environment. *International Journal on Advanced Science, Engineering and Information Technology*. 2017. Vol. 7. № 2. P. 475-481.
4. Chougala M., Kuntoji S. Novel text independent speaker recognition using LPC based formants. *International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016*. 2016. № 7755666. P. 510-513
5. Mi L., Tao S., Wang W., Dong Q., Guan J., Liu C. English vowel identification and vowel formant discrimination by native Mandarin Chinese- and native English-speaking listeners: The effect of vowel duration dependence. *Hearing Research*. 2016. Vol. 333. P. 58-65.
6. Sorokin V.N., Tsyplikhin A.I. Speaker verification using the spectral and time parameters of voice signal. *Journal of Communications Technology and Electronics.* 2010. Vol. 10. № 2. P. 87–104.
7. Реформатский А.А. Введение в языковедение. Москва: Аспент пресс. 1996. 275 с.
8. Whalen D.H., Chen W., Tiede M.K., Nam H. Variability of articulator positions and formants across nine English vowels. *Journal of Phonetics*. 2018. Vol. 68. P. 1–14.
9. Hardy T.L.D., Boliek C.A., Wells K., Dearden C., Zalmanowitz C., Rieger J.M. Pretreatment acoustic predictors of gender, femininity, and naturalness ratings in individuals with male-to-female gender identity. *American Journal of Speech-Language Pathology*. 2016. Vol. 25, №2. P. 125-137
10. Karpov A.A., Ronzhin A.L. A universal assistive technology with multimodal input and multimedia output interfaces. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2014. Vol. 8513. №1. P. 369-378.
11. Рахманенко И.А. Программный комплекс для идентификации диктора по голосу с применением параллельных вычислений на центральном и графическом процессорах. *Доклады Томского государственного университета систем управления и радиоэлектроники*. Т.20. №1. С. 70-74
12. Goto R., Misawa K., Okada Y. Analysis of Individual Characteristics in Vowel Spectral Envelopes. *Lecture Notes in Engineering and Computer Science*. 2017. Vol. 2227. P. 113-116.

13. Lavner Y., Gath I., Rosenhouse J. The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. *Speech Communication*. 2000. Vol. 30. №1. P. 9–26.
14. Коваль, С.Л. Сборник научно-методических рекомендаций по выполнению криминалистических экспертиз звукозаписей речи. СПб.: Центр речевых технологий. 2000. 174 с.
15. Hilbert D. Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen. Chelsea Pub, 1953.

I.A. Gurakov, E.Y. Kostyuchenko, D.I. Novokhrestova, A.A. Shelupanov (Tomsk State University of Control Systems and Radioelectronics, Tomsk)

Statistical distribution of formants for phonoscopic examinations using the method of formant alignment

The aim of this work is obtaining a statistical distribution of the formant frequencies for several speakers. The received distributions was used for checking the admissibility of the assumptions made earlier and the need for their correction. It was shown that the assumption about uniform distribution of the formant frequencies is not confirmed for one phoneme and one speaker and further verification of this assumption is necessary for a number of speakers. The frequencies for aligned and fully matched areas for 6 different vowel sounds are determined within the same speaker and two different speakers. It is shown as a high probability of mismatch for the aligned section of one speaker, as the possibility of a complete coincidence of the formants for different speakers in completely phonetically identical fragments. These results can be used in the construction of the phonoscope investigation method on the basis of statistical distributions of the formant frequencies.

АЛГОРИТМИЧЕСКИЙ ПОДХОД К ОПРЕДЕЛЕНИЮ ИНТОНАЦИОННОЙ ХАРАКТЕРИСТИКИ РЕЧЕВОГО ОБРАЗЦА

Я. О. Анисимов[1], **Д. А. Кацай**[2], **Т. Д. Исупова**[3]
(*Южно-Уральский государственный университет (НИУ), Челябинск*)

Аннотация

В работе описывается решение задачи определения вопросительной интонации в речевом образце. Обсуждаются требования к речевому образцу, подход к его предварительной обработке для решения поставленной задачи. Приведено обобщенное описание математической постановки задачи и способов её решения. Описывается решение задачи с использованием метода решающих деревьев и аппарата нейронных сетей. Приводятся результаты моделирования.

Ключевые слова: анализ речи, машинное обучение, нейронные сети, решающие деревья

Введение

При изучении иностранного языка помимо освоения грамматики, лексики и фонетики важным аспектом является изучение его интонационных особенностей. В русском языке интонация может играть существенную смысловую роль в слове, фразе, предложении. Например, помимо грамматически полных вопросительных предложений таких как: «Какой сейчас час?» или «Где находится университет?» существую еще большой класс как отдельных слов, так и фраз, вопросительный характер которых задаёт только интонация произносящего и их контекст. Примеры таких слов и фраз чаще всего встречаются в повседневной речи, например: «Правда?»; «Антон дома?». При использовании систем автоматического распознавания речи выделение вопросительной интонации становится затруднительной. Более того, фраза, изъятая из контекста диалога или монолога, может быть вообще не распознана как вопросительная. Для изучающих русский язык и являющихся носителями других языков, эта грань русского языка может быть недоступна. При обучении с преподавателем обратная связь и корректировка интонации происходит в процессе интерактивного взаимодействия обучаемого и преподавателя. Для изучения русского языка в системе дистанционных курсов необходимо специализированное программное обеспечение, которое в автоматическом режиме выдает интонационную оценку речи обучающегося. Разработке алгоритма работы такой программы и посвящена данная работа [5],[8],[9].

Постановка задачи

Рассматривается задача классификации речевого фрагмента по интонационному признаку к одному из классов. Предполагается, что на вход системы классификации подается речевой фрагмент длиной не более чем T_{max} секунд. На выходе системы даётся оценка принадлежности классу. Рассматриваемая задача классификации рассматривается как бинарная, поэтому система выдаёт принадлежность к заданному классу в диапазоне $[0,1]$, который можно интерпретировать как степень приближенности к указанному классу. В рамках рассматриваемой задачи будем считать что значению "1" соответствует речевой образец с вопросительной интонацией, значение "0" - отсутствие таковой.

Рассмотрим исходные данные, необходимые для решения поставленной задачи. В качестве исходных данных используется записанный и размеченный массив речевых образцов, таким образом, чтобы был сформирован массив структур:

$$\{[x_1, x_2 \ldots x_{mi}]_i, Class_i\}, i \ (1, N),$$

где $[x_1, x_2 \ldots x_{mi}]$ - запись i-го речевого образца длинной mi во временной области;

- $Class_i$ - метка класса, которому принадлежит i-й речевой образец;
- N - количество речевых образцов.

[1] Инженер-исследователь.
[2] Кандидат технических наук, доцент.
[3] Аспирант.

Рассмотрим более подробно требования к речевым образцам. Для простоты будем считать, что все образцы записаны с одинаковой частотой дискретизации. Рассмотрим более подробно вопрос о длине речевого образца. Так как все образцы уникальны, то длина образцов будет различна. Отсюда следует проблема временного выравнивания образцов. Для этого вычисляется величина T_{max}, которая соответствует длине максимального по длительности речевого фрагмента. Для выравнивания речевых образцов они дополняются «слева» нулевыми значениями фрагментов до значения T_{max}.

Решение задачи

В рамках теории машинного обучения для задач классификации могут быть использованы следующие алгоритмы:

- нейронные сети;
- решающее дерево, решающий лес;
- Байесовский классификатор;
- линейный разделитель.

Каждый из описанных методов обладает своими достоинствами и недостатками. Например, решающее дерево обладает большей информативностью для эксперта, по сравнению с другими алгоритмами; нейронные сети могут выделить не очевидные закономерности; линейный классификатор прост в реализации, но может потребовать более сложной комбинацией признаков.

Как показывает практика наибольший прирост даёт ансамблирование различных алгоритмов. В рамках работы будет рассмотрено 2 алгоритма, которые могут быть ансамблированы: нейронные сети и решающие деревья.

Решение задачи с использованием решающих деревьев.

Первые работы по использованию решающих деревьев для анализа данных появились в 60-х годах 20 века. Несмотря на свою интерпретируемость и высокую выразительную способность, деревья крайне трудны для оптимизации из-за своей дискретной структуры — дерево нельзя продифференцировать по параметрам и найти с помощью градиентного спуска экстремум. Более того, даже число параметров у них не является постоянным и может меняться в зависимости от глубины и выбора критериев дробления. Из-за этого все методы построения решающих деревьев являются "жадными" и эвристичными [1].

Рассмотрим подготовку данных для решения поставленной задачи описанным методом. Подавать массив значение на прямую в дерево не целесообразно, т. к. решающие деревья сами по себе не выделяют каких-либо признаков входных величин. Выделение признаков — задача разработчика. При использовании речевого сигнала можно выделить следующие признаки:

- хромаграмма из формы волны или спектрограммы мощности;
- Q-хромаграмма;
- вариант цветности «Chroma Energy Normalized» (CENS);
- мел-частотные кепстральные коэффициенты;
- среднеквадратичная (RMS) энергия для каждого кадра;
- спектральный центроид;
- спектральная ширина полосы p-го порядка;
- спектральная плотность;
- спектральный контраст;

Применяя указанные признаки к фрагменту получим их массив:

$$F = [feature_1, feature_2 \dots feature_9].$$

Далее распределяем признаки по временной шкале. Ввиду того, что все образцы после выравнивания имеют одинаковую длину, можно получить все признаки для отдельного участка образца. Длина участка выбирается фиксированной, чтобы разбить весь образец на равные интервалы. Таким образом, мы получим итоговый массив:

$$F_I = \left[feature_1^{(1)}, feature_2^{(1)} \dots feature_9, \dots feature_{9,}^{(2)} feature_1^{(K)}, feature_2^{(K)} \dots feature_9^{(K)}\right]$$

где K — количество участков речевого образца.

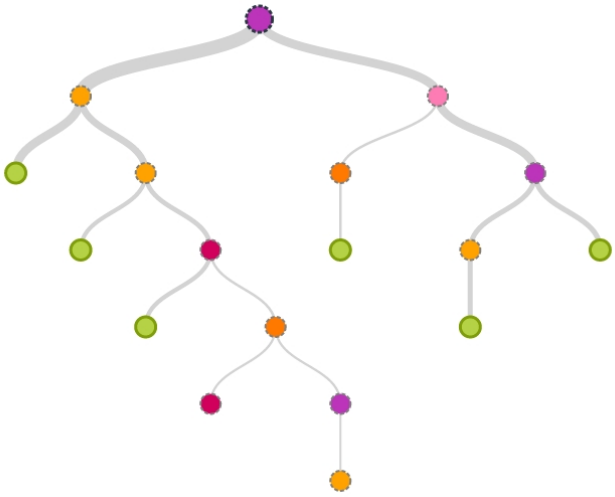


Рис. 1. Пример решающего дерева

Опишем базовый "жадный" алгоритм построения бинарного решающего дерева. Начнем со всей обучающей выборки X и найдем наилучшее ее разбиение на две части $R_1(j,t)=\{F\ F_j<\Theta\}$ и $R_2(j,t)=\{F\ F_j>\Theta\}$ с точки зрения заранее заданного функционала качества $Q(F,j,\Theta)$. Найдя наилучшие значения j и Θ, создадим корневую вершину дерева, поставив ей в соответствие предикат $R_1(j,t)=\{F\ F_j<\Theta\}$. Объекты разобьются на две части — одни попадут в левое поддерево, другие в правое. Для каждой из этих подвыборок рекурсивно повторим процедуру, построив дочерние вершины для корневой, и так далее. В каждой вершине проверяем, не выполнилось ли некоторое условие останова — и если выполнилось, то прекращаем рекурсию и объявляем эту вершину листом. Когда дерево построено, каждому листу ставится в соответствие ответ. В случае с классификацией это может быть класс, к которому относится больше всего объектов в листе, или вектор вероятностей (например, вероятность класса может быть равна доле его объектов в листе). В качестве численного значения оценки может быть среднее арифметическое значение, среднее структурное значение (медиана) или другая функция от целевых переменных объектов в листе. Выбор конкретной функции зависит от функционала качества в исходной задаче

При построении дерева необходимо задать функционал качества, на основе которого осуществляется разбиение выборки на каждом шаге. Обозначим через R_m множество объектов, попавших в вершину, разбиваемую на данном шаге, а через R_l и R_r — объекты, попадающие в левое и правое поддерево соответственно при заданном предикате. Будем использовать функционалы следующего вида: $Q(R_m,j,s)=H(R_m)+card(R_l)card(R_m)H(R_l)-card(R_l)card(R_m)H(R_r)$. Здесь $H(R)$ — это критерий информативности, который оценивает качество распределения целевой переменной среди объектов множества R. Чем меньше разнообразие множества значений целевой переменной, тем меньше должно быть значение критерия информативности $H(R)$. Функционал качества $Q(R_m,j,s)$ при этом будем максимизировать. Как уже обсуждалось выше, дерево будет выдавать константу — класс в каждом своем листе. Исходя из этого критерия, будем оценивать качество множества объектов R значением целевых переменных: насколько хорошо они предсказывают классы [2], [3].

Решение задачи с использованием нейронных сетей.

Нейронные сети хорошо себя зарекомендовали для решения различных задач: аппроксимации, фильтрации, прогнозирования и классификации. До 2010-х годов применение нейронных сетей для неструктурированных данных вызывало ряд трудностей. Одним из примеров указанных сложностей является задача классификации на массиве данных ImageNet. Каждое изображение аннотировано на наличие или отсутствие объекта в изображении, например: «на этом изображении есть тигры» или «на этом изображении нет тигров». Аннотацию уровня объекта

предоставляет ограничивающий прямоугольник вокруг видимой части указанного объекта. ImageNet использует вариант широкой схемы WordNet для кластеризации объектов, дополненных 120 категориями пород собак, чтобы продемонстрировать мелкозернистую классификацию [4].

В 2012 году на конкурсе по классификации изображений ImageNet был совершен прорыв. Группой исследователей под руководством Крижевского [6] была представлена нейронная сеть AlexNet, которая достигла ошибки 15,1%. Для сравнения, алгоритмы которые использовались до этого события достигали ошибки не менее 26,1%. Помимо соревнования ImageNet различные архитектуры нейронных сетей: ZF Net, VGG Net, Inception, ResNet, показали отличное качество распознавания и для других массивов данных (CIFAR-10,CIFAR-100).

Т.к. задача классификации изображений имеет устоявшиеся модели которые дают минимальную величину ошибки, то сведем задачу классификации речевого фрагмента к задаче классификации изображения. Условимся, что под математической моделью изображения, в общем случае, будем предполагать тензор размера [3·Ширина·Высота]. В случае, когда речь будет идти о монохромном изображении, входным параметром будет матрица размером [1·Ширина·Высота].

Для преобразования временного ряда в «изображение» будем применять операции получения спектрограммы и мелспектрограммы из речевого образца, которые можно вычислять в реальном времени. Таким образом, можно получить визуальный отпечаток речевого сигнала.

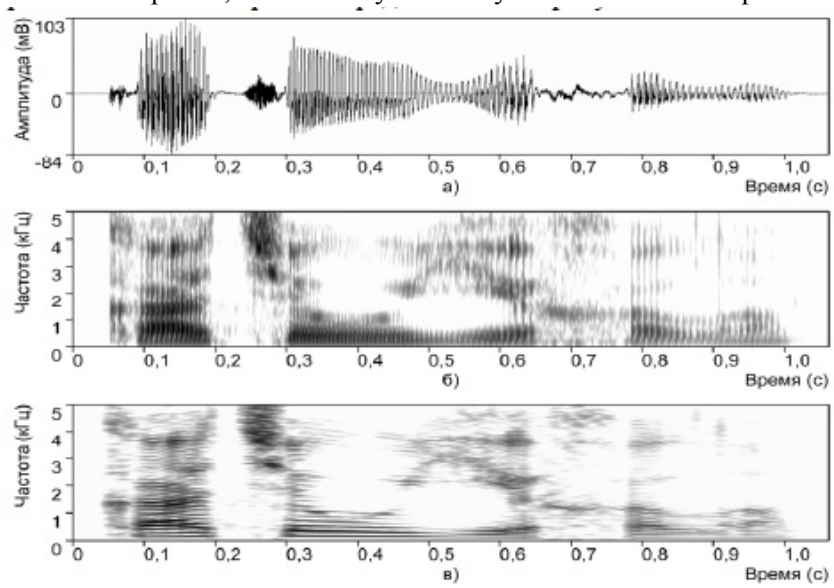


Рис. 2. Спектрограммы речевого фрагмента: а) осциллограмма; б) широкополосная спектрограмма; в) узкополосная спектрограмма

Достоинство архитектуры нейронной сети AlexNet, состоит в использовании сверточных слоев в ее составе. Сети, содержащие такой слой в своем составе, получили название сверточные. Обычно в подобных нейронных сетях помимо сверточного слоя присутствует слой субдискретизации и полносвязный слой. Теория использования подобных слоев в составе нейронных сетей была предложена Лекун [7].

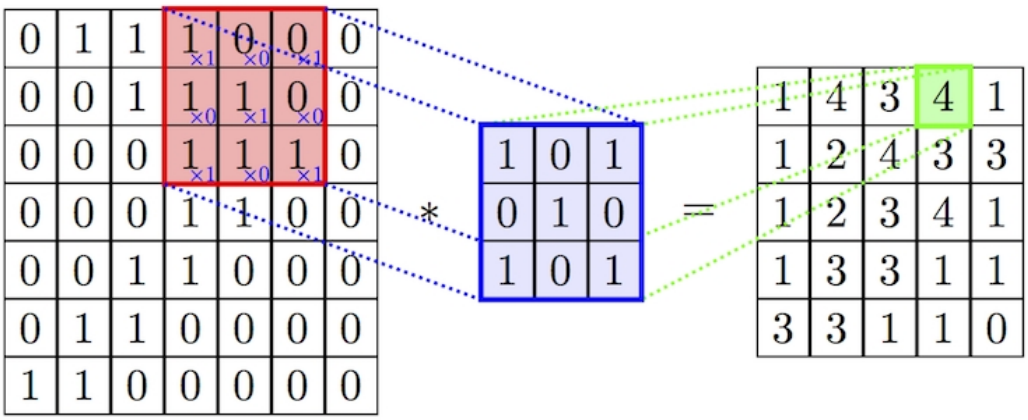


Рис. 3. Работа сверточного слоя нейронной сети

В сверточных нейронных сетях слои свёртки и субдискретизации состоят из нескольких «уровней» нейронов, называемых картами признаков. Каждый нейрон такого слоя соединён с небольшим участком предыдущего слоя, называемым рецептивным полем. В случае изображения, карта признаков является двумерным массивом нейронов, или просто матрицей. Другие измерения могут быть использованы, если на входе принимается другой вид данных, например, аудио данные (одномерный массив) или объёмные данные (трёхмерный массив). В слое свёртки каждой карте признаков соответствует одно ядро свёртки, также называемое фильтром. Каждый нейрон в качестве своего выходного значения содержит результат операции свёртки или взаимной корреляции со своим рецептивным слоем.

Стоит заметить, что эти две операции в контексте обучения свёрточных нейронных сетей взаимозаменяемы, вследствие чего во многих программных реализациях операция “свёртки” на самом деле является операцией взаимной корреляции.

Так как ядро свёртки для каждой карты признаков одно, то это позволяет нейронной сети "научиться" выделять признаки вне зависимости от их расположения во входном изображении и также приводит к значительному уменьшению числа параметров сети.

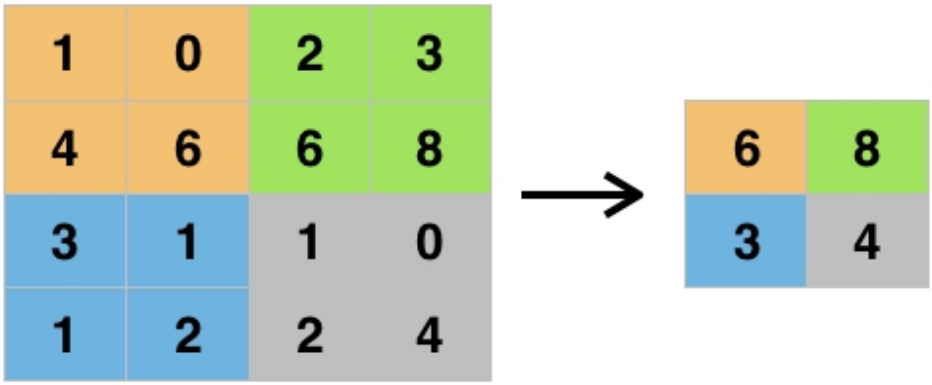


Рис. 4. Принцип работы слоя субдискретизации нейронной сети

Слой субдискретизации осуществляет уплотнение карт признаков предыдущего слоя и не изменяет количества карт. Каждая карта признаков слоя соединена с соответствующей картой признаков предыдущего слоя, каждый нейрон выполняет «сжатие» своего рецептивного поля посредством какой-либо функции.

Наиболее популярными видами этого слоя являются Max Pooling (из рецептивного слоя выбирается максимальное значение), Average Pooling (выбирается среднее значение) и L2 Pooling (выбирается норма L2). С помощью слоя субдискретизации достигается устойчивость к небольшим сдвигам входного изображения, а также уменьшается размерность последующих слоёв.

Полносвязный слой — обычный скрытый слой многослойного перцептрона, соединённый со всеми нейронами предыдущего слоя. Таким образом, на вход свёрточной нейронной сети подаётся изображения, а на выходе получается класс, к которому принадлежит изображение.

Результаты моделирования

Рассмотрим использование описанных методов для решения описанной задачи. Важным вопросом является получение качественной обучающей выборки. Возможны два варианта формирования выборки:

- самостоятельная запись обучающей выборки, с учётом всех требований в исходной постановке задачи;
- поиск и использование заранее размеченной выборки.

Для отработки предложенных алгоритмов был выбран вариант с использованием готовой выборки. Такой подход обладает рядом недостатков. Обычно разметки подобных выборок создаются на основе свободных записей аудиокниг. Такие записи создаются профессиональными дикторами и интонационный характер их речи существенно отличается от повседневной интонации. Также существует ограничение на текст, с которого происходит начитка образцов. Дело в том, что в художественных произведениях присутствует существенный дисбаланс между во-

просительными и иными фразами. Также открытым остается вопрос о контекстуальном значении той или иной фразы. Указанные недостатки можно считать естественным шумом обучающей выборки. Однако, несмотря на указанные недостатки, использование подобных обучающих выборок можно использовать в качестве первого шага для отработки алгоритмов выявления интонационных признаков в речевых образцах.

В качестве обучающей выборки выбраны данные "The M-AILABS Speech Dataset" из исследовательского проекта, предназначенного для создания систем распознавания речи на 9 языках: английский, немецкий, русский, украинский, турецкий и других. Для русского языка параметры выборки следующие:

- размер выборки - 3.6 Гигабайта;
- общая длительность выборки 47 часов;
- количество голосов — 3 (2 мужских и один женский);
- общее количество образцов — 20500.

Проведенный анализ обучающей выборки показал следующее соотношение между вопросительными и иными фразами: 4670 вопросительных и 15830 иных предложений. Необходимость подобный дисбаланса для обучения предложенных алгоритмов является вопрос открытым. С одной стороны в обычной речи имеются подобные соотношения между обсуждаемыми классами, и целесообразно было бы сохранить всю выборку. С другой стороны в исходной постановке задачи предполагается управляемое взаимодействие между человеком и разрабатываемой системой и логично свести выборку к построению двух сбалансированных классов речевых образцов.

Для реализации предложенных алгоритмов были разработаны алгоритмы на языке Python с использованием следующих библиотек: scikit-learn, Keras, Kapre. Из исходной выборки была сформирована сбалансированная обучающая выборка с новой разметкой данных.

При реализации алгоритма решающих деревьев были проварьированы следующие параметры: количество участков речевого образца, глубина решающего дерева. В результате экспериментов были получены следующие результаты.

Таблица 1

Результаты моделирования алгоритма решающего дерева

Глубина дерева	4	5	6
Точность оценки	57%	65%	58%

Для реализации алгоритма на основе нейронных сетей была создана сеть, архитектура которой изображена на рисунке.

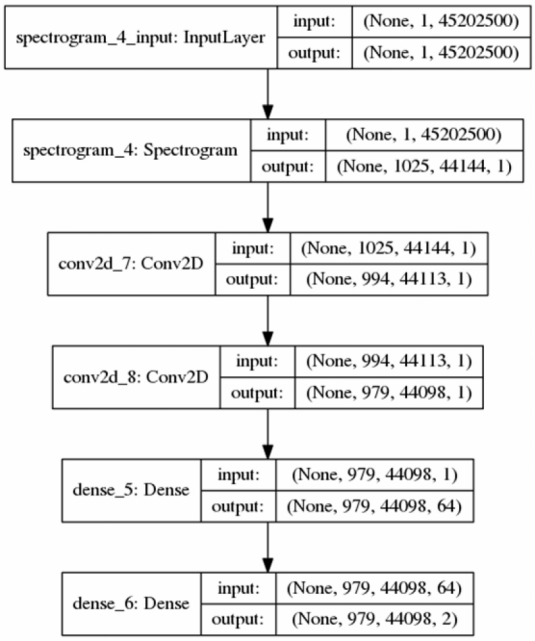


Рис. 5. Архитектура разработанной сети

После обучения точность работы нейронной сети составила 56 процентов. Такую низкую точность можно объяснить несколькими причинами. Во-первых сама исходная база для обучения предназначена больше для задач преобразования речи в текст, а не для работы с интонационной характеристикой речевого образца. Вторая, возможная, причина может скрываться в предварительной обработке речевого образца, то есть выбора входных данных для нейронной сети. Также остается открытым вопрос об оптимальной архитектуре сети. Одним из методов повышения точности при имеющихся исходных данных видится ансамблирование указанных алгоритмов, либо перестроение нейронной сети, а конкретно полносвязанных слоёв, с детерминированной на стохастическую.

ЛИТЕРАТУРА

1. Anyanwu M. N., Shiva S. G. Comparative analysis of serial decision tree classification algorithms //International Journal of Computer Science and Security. – 2009. – Т. 3. – №. 3. – С. 230-240.
2. Brijain M. et al. A survey on decision tree algorithm for classification. – 2014.
3. Chauhan H., Chauhan A. Implementation of decision tree algorithm c4. 5 //International Journal of Scientific and Research Publications. – 2013. – Т. 3. – №. 10.
4. Deng J. et al. Imagenet: A large-scale hierarchical image database //Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. – Ieee, 2009. – С. 248-255.
5. D. Katsay, E. Doronina, Y. Kazakova, E. Kharchenko, T. Isupova (2017) ANALYSIS OF SPEECH SAMPLES AS A MEANS OF INTENSIFICATION OF TEACHING PHONETICS OF A FOREIGN LANGUAGE, ICERI2017 Proceedings, pp. 2771-2778.
6. Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks //Advances in neural information processing systems. – 2012. – С. 1097-1105.
7. LeCun, Y., Kavukcuoglu, K., Farabet, C. Convolutional networks and applications in vision. In Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS) (2010). IEEE, 253–256.
8. Д.А. Кацай, Т.Д. Исупова, Е.В. Харченко. «Алгоритм анализа речевых образцов иностранных студентов, изучающих русский язык». / Международная научно-практическая интернет-конференция «Актуальные вопросы описания и преподавания русского языка как иностранного/неродного» (Москва, 27 ноября – 1 декабря 2017 г.): Сборник материалов / Под общ. ред. Н.В. Кулибиной. – М., 2018. – 1074 с.: ил. [Электронное издание]. ISBN 978-5-98269-173-6 https://confrki.pushkininstitute.ru/content/Сборник.pdf С.43-48.
9. Д.А. Кацай, Т.Д. Исупова, Е.В. Харченко, А.Д. Бирюков, Д.С. Филиппов. «Структура тренажёра для совершенствования речи иностранных студентов, изучающих русский язык». / Международная научно-практическая интернет-конференция «Актуальные вопросы описания и преподавания русского языка как иностранного/неродного» (Москва, 27 ноября – 1 декабря 2017 г.): Сборник материалов / Под общ. ред. Н.В. Кулибиной. – М., 2018. – 1074 с.: ил. [Электронное издание] ISBN 978-5-98269-173-6 https://confrki.pushkininstitute.ru/content/Сборник.pdf С. 492-495.

Y.O. Anisimov, D.A. Katsai, T.D. Isupov (South Ural State University, Chelyabinsk)

Algorithmic approach to determination of intonative characteristics of the speech sample

When teaching a foreign language an important aspect is teaching the correct pronunciation of words or phrases. In full-time education, a teacher, working in pairs with a student or group of students, acts as an observer, evaluating the correctness of the pronunciation of the student, and in the role of "control device," correcting and explaining exactly how to change the mechanics of pronunciation. With online training, in the form of MOOC, it is difficult to single out a specialist to perform such functions. Therefore, there is a need to develop a software package for assessing the correctness of the pronunciation of the learner.

Traditionally, the problem of speech analysis uses approaches based on frequency and cepstral analysis of the sound sample and its comparison with the standard. When using this approach, it is required to create a qualitative base of standards for calculating the cepstral coefficients for a particular sound. Also to the lack of this approach is the comparative complexity when working with high-grade mods of whole phrases: such as intonation, accent, content. However, the use of cepstral coefficients in combination with other characteristics of the sound sample makes it possible to construct an algorithm or an ensemble of algorithms for solving various problems.

An alternative is to use neural networks to solve the tasks. Neural networks have shown themselves well when solving problems of classification of non-manufactured data, for example, images. Using docked and fully connected layers, it is possible to form a structure that would perform the task of classifying sound. In this case, the sound will be presented not in the form of a time sequence, but in the form of a one-stage snapshot of the entire phrase.

The report describes the solution of the task of highlighting interrogative intonation in the spoken tuition phrase. The problems of creating a training sample are discussed. Several variants of the architecture of neural networks, capable of solving the task, as well as their learning algorithms, are described. The results of the simulation are presented.

ИССЛЕДОВАНИЕ МЕТОДОВ ПОСТРОЕНИЯ ИНТЕГРАЛЬНЫХ СИСТЕМ РАСПОЗНАВАНИЯ РУССКОЙ РЕЧИ

Н. М. Марковников[1], **И. С. Кипяткова**[2]
(*Федеральное государственное бюджетное учреждение науки
«Санкт-Петербургский институт информатики и автоматизации Российской академии наук»,
С.-Петербург, 14 линия, дом 39, niklemark@gmail.com, kipyatkova@iias.spb.su*)

Аннотация

Интегральные системы распознавания речи, использующие глубокие нейронные сети (DNN), показывают хорошие и перспективные результаты. В данной работе предлагается применение коннекционной временной классификации (CTC) и механизма-внимания для построения интегральных моделей распознавания русской слитной речи. В экспериментах были использованы различные архитектуры нейронных сетей, такие как: долгая кратковременная память (LSTM), двунаправленная LSTM и остаточная сверточная сеть. Полученные модели продемонстрировали точность распознавания немного хуже, чем гибридные модели, но данные модели могут работать без использования большой языковой модели и показывают высокую среднюю скорость обучения и декодирования, что может быть важным в реальных системах распознавания.

Ключевые слова: интегральные модели, глубокое обучения, русская речи, распознавание речи, нейронные сети.

Введение

Цель автоматического распознавания речи — преобразование звукового сигнала в последовательность слов. Стандартные системы автоматического распознавания речи разбивают задачу на три шага: выделение признаков, акустическое моделирование и декодирование последовательности. Данные системы обычно строятся с использованием скрытых марковских моделей в качестве акустических моделей и *n*-граммных моделей – в качестве модели языка. Такие системы показывают приемлемую точность распознавания, но они состоят из двух частей, настраиваемых независимо. Также стандартные системы распознавания требуют большого объема памяти и вычислительной мощности, что не позволяет использовать подобные системы локально на мобильных устройствах и требует удаленных вычислений на серверах.

Интегральные системы (англ. *end-to-end*) распознавания речи позволяют использовать глубокие нейронные сети, обычно только одну, которые обучаются глобально с использованием градиентного спуска и одной функцией потерь, что позволяет преобразовывать входящие последовательности акустических признаков сразу в выходную последовательность слов. Такой подход часто демонстрирует лучшую производительность с точки зрения скорости работы и точности распознавания. Интегральные модели требует малого объема памяти, но большего количества данных для обучения моделей.

В последнее время получили распространение модели, использующие коннекционную временную классификацию (Connectionist Temporal Classification; CTC) в качестве функции потерь нейронной сети. В работе был исследован метод реализации интегральной системы распознавания речи с использованием коннекционной временной классификации.

Еще один тип моделей, используемый для построения интегральных систем распознавания речи – это шифратор-дешифратор (англ. *encoder-decoder*) модели, которые показывают результаты лучше, чем гибридные системы распознавания речи и модели на основе коннекционной временной классификации.

Целью данной работы была реализация интегральной системы распознавания русской речи. В работе был исследован метод реализации интегральной системы распознавания речи с использованием шифратор-дешифратор модели, основанной на механизме внимания, и модели, использующей коннекционную временную классификацию. Были проведены эксперименты по сравнению качества и скорости распознавания речи, достигнутых с применением созданных моделей и базовой системой распознавания речи, в качестве которой была использована система с гибридными акустическими моделями на основе скрытых марковских моделей. Качество

[1] Магистрант, программист.
[2] Кандидат технических наук, старший научный сотрудник.

распознавания речи измерялось с помощью двух метрик: ошибка распознавания символов (англ. CER) и ошибка распознавания слов (англ. WER).

Интегральные модели распознавания русской речи

Аналитический обзор подходов к построению интегральных систем распознавания речи. В [1] и [2] была представлена реализация интегральной системы с использованием инструментария Eesen [1], где декодирование CTC моделей происходило с помощью взвешенных конечных преобразователей (англ. WFST) [3]. Каждый компонент системы: CTC метки (T), словарь (L) и языковая модель (G) — преобразовывались в один граф поиска следующим образом:

$$TLG = T \circ \min(\det(L \circ G)),$$

где min означает минимизацию, det — детерминацию, и $\circ$ — композицию. Так, TLG строит соответствие между последовательностью CTC меток и словами. Это позволяет производить эффективный поиск с помощью, например, библиотеки OpenFST [4]. В [2] были использованы двунаправленные LSTM-сети для распознавания сербской речи. При этом был получен WER, равный 14,68%, что является не самым хорошим результатом, хотя CER оказалась довольно маленькой — 3,68%.

В [5] было проведено исследование интегральных систем с использованием CTC. Было показано, что CTC модель может хорошо работать и без языковых моделей и словаря. В качестве обучающего речевого корпуса был использован корпус [6], составленный из аудиодорожек YouTube видео, общей длительностью более 650 часов. А в качестве тестового корпуса были взяты аудиодорожки из Google Preferred [7] видео, общей длительностью 25 часов. Лучшая WER без использования модели языка была равна 13,9%, а с моделью языка — 13,4%.

В [8] также была предложена интеграция модели, основанной на механизме внимания, и языковой модели. Для того чтобы построить модель языка, основанную на символах, из модели, построенной на словах, использовались WFST. При декодировании запускался поиск выхода, который минимизировал некоторый функционал, комбинирующий шифратор-дешифратор модель и модель языка. В итоге на речевом корпусе WSJ были получены следующие результаты: WER, равная 11,3%, и CER, равная 4,8%.

В [9] независимо была предложена похожая модель, основанная на механизме внимания, названная «Listen, Attend and Spell» (LAS). Шифратор представлял собой BLSTM пирамидальной структуры, а дешифратор использовал LSTM. Также полученная модель после декодирования пересчитывалась с помощью модели языка. Так на речевом корпусе Google Voice Search была получена WER, равная 10,3%.

Как можно видеть, интегральные модели показывают приемлемые результаты как с использованием, так и без языковых моделей для языков со строгим порядком слов (например, английский).

Русский язык характеризуется высокой степенью грамматической свободы и сложностью словообразования. Поэтому возникает необходимость использования внешних языковых моделей для увеличения точности распознавания. В любом случае, пока что не было получено интегральных моделей для русского языка, поэтому было решено провести данное исследование.

Коннекционная временная классификация. Коннекционная временная классификация (Connectionist Temporal Classification; CTC) [10] — это функция, которая позволяет рекуррентным нейронным сетям обучаться для распознавания последовательности слов без начального выравнивания входных и выходных последовательностей.

Опишем подход, в котором CTC-функция используется в качестве функции потерь для обучения нейронной сети. Выходной слой нейронной сети содержит по одному блоку для каждого символа выходной последовательности и еще один для дополнительного символа «пропуск» («blank»), соответствующего пустому выходному символу. Выходной вектор w_m нормализуется с помощью softmax [11] функции, которая интерпретируется как вероятность появления символа (или «пропуска») с индексом k в момент времени m:

$$P(k, m \mid x) = \frac{\exp(w_m^k)}{\sum_{k'=0}^{|w_m|} \exp(w_m^{k'})},$$

где x означает последовательность входных признаков длины T, w_m^k — k-ый элемент w_m, а $|w_m|$ — длина слова w_m. Пусть, α — последовательность из индексов «пропусков» и символов длины T для выравнивания. Вероятность $P(\alpha \mid x)$ можно представить как произведение вероятностей появления символов в каждый момент времени:

$$P(\alpha \mid x) = \prod_t^T P(\alpha_t, t \mid x).$$

Для данной выходной последовательности существует столько возможных выравниваний, сколько способов расставить «пропуски» между символами. Обозначим, что B — оператор, который удаляет сначала все повторы, а затем — «пропуски». Так, полная вероятность выходной последовательности w равна сумме вероятностей всех возможных соответствующих выравниваний:

$$P(w \mid x) = \sum_{\alpha \in \mathrm{B}^{-1}(w)} P(\alpha \mid x),$$

где B^{-1} — оператор, обратный к B. Эта сумма может быть вычислена с помощью метода динамического программирования [10].

Пусть w^* — целевая последовательность слов, тогда нейронная сеть может быть обучена минимизировать CTC функцию:

$$CTC(x) = -\log P(w^* \mid x).$$

В [10] был предложен метод декодирования интегральных CTC-моделей. Метод основывается на предположении, что наиболее вероятное выравнивание соответствует наиболее вероятной выходной последовательности:

$$\arg\max_w P(w \mid x) \approx \mathrm{B}(\alpha^*),$$

где $\alpha^* = \arg\max_\alpha P(\alpha \mid x)$.

В [12] был предложен метод декодирования, использующий алгоритм лучевого поиска (beam search algorithm), который также позволяет интегрировать языковую модель.

Шифратор-дешифратор модель с механизмом внимания. Шифратор-дешифратор (Encoder-Decoder) модели часто используются для задач, где длины входной и выходной последовательностей являются переменными [13, 14]. Шифратор (Encoder) — это нейронная сеть, которая трансформирует вход $x = (x_1, \ldots, x_{L'})$ в некоторое промежуточное представление $h = (h_1, \ldots, h_L)$, выделяет признаки. Дешифратор (Decoder) — это обычно рекуррентная нейронная сеть, которая использует это промежуточное представление для генерации выходных последовательностей.

На i-ом шаге дешифратор генерирует выход y_i, фокусируясь на определенных элементах h:

$$\alpha_i = \mathrm{Attend}(s_{i-1}, \alpha_{i-1}, h)$$

$$g_i = \sum_{j=1}^{L} \alpha_{i,j} h_j$$

$$y_i = \mathrm{Generate}(s_{i-1}, g_i),$$

где s_{i-1} — $(i-1)$-е состояние рекуррентной сети, которое называется Generator, $\alpha_i \in \mathbb{R}^L$ — вектор весов внимания (англ. attention weights), которые также часто называются выравниванием [13]. В [14] g_i было названо «проблеск» (англ. glimpse). Шаг завершается вычислением нового состояния генератора:

$$s_i = \mathrm{Recurrency}(s_{i-1}, g_i, y_i).$$

Условия проведения экспериментов. В данной работе был использован обучающий речевой корпус, собранный в СПИИРАН [15]. Корпус состоит из трех частей:

1. записи 50 дикторов, носителей русского языка, где каждый диктор произнес 327 фраз;
2. записи 55 дикторов, носителей русского языка, где каждый диктор произнес 105 фраз;
3. аудио-часть корпуса HAVRUS [16], где 20 дикторов (10 мужчин и 10 женщин) произнесли по 200 фраз.

Общая продолжительность записей, входящих в корпус, более 30 часов.

Для тестирования системы использовался речевой корпус из 500 фраз, произнесенных 5 дикторами. Фразы были взяты из материалов российской онлайн-газеты «Fontanka.ru[1]».

Языковая модель была обучена на данных российских новостных сайтов [17]. Размер обучающего речевого корпуса составил примерно 300 миллионов словоупотреблений. В качестве модели была использована n-грамная модель со сглаживанием Кнесера-Нея [18] с $n = 2$ и $n = 3$. Размер словаря – 150 тыс. словоформ.

Базовое решение. Базовым решением являлась гибридная модель с использованием скрытых марковских моделей, которая была реализована с помощью инструментариев Kaldi [19] и CNTK[2], аналогично работе [20]. При декодировании использовалась биграммная языковая модель. Лучшие результаты, полученные с использованием двунаправленной LSTM, ResNet [21] и рекуррентной сверточной сети (RCNN) [22], представлены в Таблице 1:

Таблица 1

Лучшие результаты базовых решений в показателях WER, средней скорости обучения (признаков в секунду) и декодирования (фраз в секунду)

Модель	WER, %	Декодирование	Обучение
Двунаправленная LSTM	23.08	0.211	450.7
ResNet	22.17	0.105	121.4
RCNN	22.56	0.162	325.1
RCNN+ResNet+BLSTM	22.07	0.197	502.3

Реализация интегральной системы распознавания речи. В начале была реализована простая система распознавания речи с использованием коннекционной временной классификации в качестве функции потерь с помощью библиотеки TensorFlow. Такая система не позволяет использовать языковую модель, но потребляет малое количество памяти.

Затем был использован инструментарий Eesen [1], в котором Kaldi использует реализацию нейросетевых моделей TensorFlow[3]. Такая система позволяет использовать языковые модели в формате Kaldi без дополнительных преобразований.

Для проведения экспериментов с механизмами внимания была использована библиотека Tensor2Tensor[4]. Данная библиотека реализует общий подход к построению моделей для преобразования одних последовательностей в другие последовательности, в частности для распознавания речи.

Эксперименты были проведены с использованием NVIDIA GeForce GT 730M с 2 Гб памяти, библиотеки CUDA и CPU с 16 Гб памяти и 4 ядрами.

Результаты экспериментов по распознаванию слитной русской речи, полученные с помощью моделей на основе коннекционной временной классификации. В экспериментах было использовано два типа признаков, выделенных их аудиозаписей с 1 каналом и с частотой равной 16000 МГц:

1. Мел-кепстральные коэффициенты (англ. MFCC) [23] с размерностью равной 13, которые были выделены с шириной окна, равной 25 мс, и шагом окна, равным 1 мс, вместе с 3 дополнительными признаками, являющимися производными первого и второго порядка.
2. 40-размерные признаки, полученные с помощью полосового фильтра (англ. filterbank) [24].

Обучающий корпус был разделен на две части: тренировочное множество (95%) и множество для кросс-валидации (5%).

Были использованы несколько типов нейронных сетей: обычный многослойный перцептрон (англ. MLP), LSTM, двунаправленная LSTM (англ. BLSTM), ResNet. Настройки данных сетей, которые показали лучшие результаты без использования языковой модели, были использованы следующие:

- Обычный перцептрон содержал 4 скрытых слоя с 512 узлами, ReLU функцией активации, начальным коэффициентом скорости обучения равным 0,007.

[1] Онлайн-газета «Fontanka.ru». URL: http://www.fontanka.ru.
[2] Инструментарий для глубокого обучения CNTK. URL: https://docs.microsoft.com/ru-ru/cognitive-toolkit.
[3] Инструментарий для глубокого обучения Tensorflow. URL: https://www.tensorflow.org.
[4] Библиотека на основе инструментария TensorFlow. URL: https://github.com/tensorflow/tensor2tensor.

- LSTM с 4 слоями и 512 блоками в каждом слое с коэффициентом регуляризации равным 0.5, начальным коэффициентом скорости обучения равным 0,001.
- LSTM со сверточными слоями в начале LSTM сети (Conv+LSTM), позволяющими упростить входные признаки. Сеть содержала один двумерный сверточный слой с 8 фильтрами и ReLU функцией активации. Затем LSTM сеть содержала 4 слоя с 512 блоками в каждом слое.
- Архитектура ResNet сети представлена на Рисунке 1. Она содержала 9 остаточных блоков с батч-нормализацией [25].

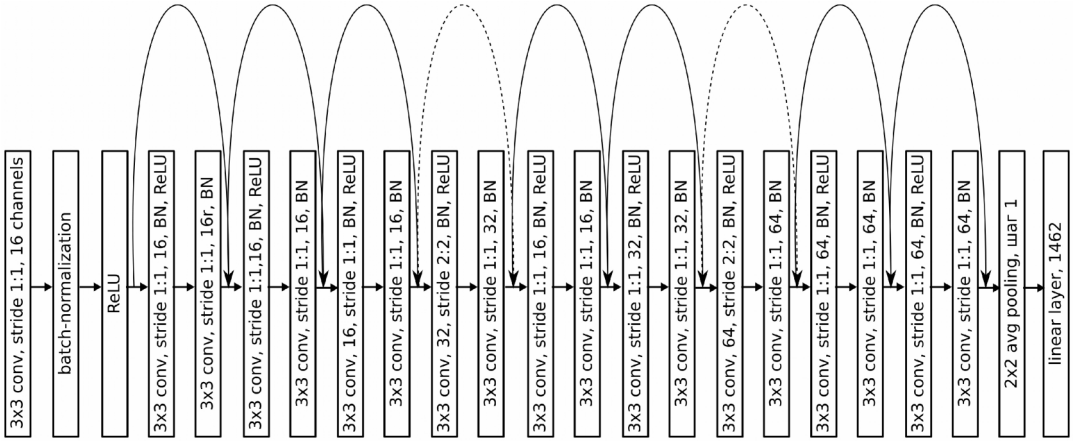


Рис. 1. Схема архитектуры, использованной в экспериментах с ResNet сетью

Для обучения сетей был использован Momentum алгоритм с коэффициентом равным 0.9. Для построения взвешенных конечным автоматов для декодирования с помощью инструментария Eesen была использована библиотека OpenFST. Аналогично работе [26] каждый компонент системы: CTC метки (T), словарь (L) и языковая модель (G) — преобразовывались в один граф поиска следующим образом:

$$TLG = T \circ \min(\det(L \circ G)),$$

где min означает минимизацию, det — детерминацию, и $\circ$ — композицию.

Использование триграммной языковой модели было затруднительно ввиду недостатка необходимой памяти и вычислительных мощностей, поэтому триграммная модель использовалась не во всех экспериментах.

Результаты экспериментов, полученные с применением моделей на основе коннекционной временной классификации, представлены в Таблице 2.

Таблица 2

Результаты, полученные в экспериментах с моделями на основе коннекционной временной классификации

Модель	CER, %	WER, %	Декодирование	Обучение
Модели без использования языковых моделей, MFCC признаки, реализация (1)				
MLP	55.42	71.64	0.252	96.7
LSTM	38.58	52.47	0.266	304.9
Conv+LSTM	36.92	49.23	0.278	315.2
BLSTM	36.73	48.86	0.282	391.7
ResNet	35.69	48.24	0.267	142.6
Модели с использованием биграммной языковой модели, MFCC признаки, реализация (2)				
MLP	48.49	62.04	0.174	125.3
LSTM	26.12	38.71	0.181	402.8
Conv+LSTM	25.77	36.93	0.193	391.1
BLSTM	22.98	35.21	0.102	407.2
ResNet	22.35	34.96	0.173	293.2
Модели с использованием триграммной языковой модели, filterbank признаки, реализация (2)				
MLP	26.57	37.19	0.098	104.7
BLSTM	15.79	26.17	0.026	381.5
ResNet	14.96	25.53	0.083	201.9

Результаты экспериментов по распознаванию слитной русской речи, полученные с помощью шифратор-дешифратор моделям с механизмом внимания. Для обучения шифратор-дешифратор моделей с механизмом внимания были использованы MFCC признаки, такие же как для обучения моделей с использованием коннекционной временной классификации. В данных экспериментах не была использована интеграция с языковыми моделями. Полученные результаты представлены в Таблице 3.

Таблица 3

Результаты, полученные с помощью шифратор-дешифратор моделей с использованием механизма внимания

Модель	CER, %	WER, %	Декодирование	Обучение
LSTM	19.15	28.47	0.279	385.2
BLSTM	19.08	27.83	0.285	401.8

Были проведены эксперименты с LSTM и двунаправленными LSTM моделями. Модели содержали 4 слоя по 128 блоков с начальным уменьшающимся коэффициентом регуляризации равным 0,9 в шифраторе. В качестве дешифратора использовалась LSTM сеть с механизмом внимания.

Был использован механизм внимания по Богданову [27] со 128 скрытыми узлами в скрытых слоях. Размер батча был равен 36 с коэффициентом скорости обучения равным 0,05.

Был использован оптимизационный алгоритм по Адаму [28] с $\beta_1 = 0.85, \beta_2 = 0.997$ и $\epsilon = 10^{-6}$. Веса нейронных сетей были инициализированы значениями из равномерного распределения из интервала $[-1;1]$.

Анализ результатов. Как и ожидалось, модели на основе коннекционной временной классификации без интеграции с языковыми моделями показывают посредственные результаты для русской речи. Наилучший полученный результат оказался равен 14.96% в показателях CER и 25.53% в показателях WER для ResNet сети с использованием языковой модели.

Как уже говорилось во введении, русский язык характеризуется высокой степенью грамматической свободы и сложностью словообразования. Поэтому, языковая модель является важной частью системы распознавания русской речи. Но было обнаружено, что модели с использованием механизма внимания позволяют достичь приемлемых результатов и без языковых моделей. Так, двунаправленная LSTM сеть с механизмом внимания показала результат 19.08% по показателю CER и 27.83% по показателю WER.

Полученные модели не смогли превзойти базовую систему на основе гибридных моделей. Однако, полученные модели показали лучше результаты по скорости декодирования и обучения, что может быть полезно в реальных системах распознавания речи. Так, LSTM модели с механизмом внимания без использования языковых моделей показали скорость декодирования равную 0,285 последовательностей в секунду, что на 19% лучше базового решения.

Заключение

В данной работе была рассмотрена задача распознавания русской речи с использованием интеграционных моделей, таких как модели на основе коннекционной временной классификации и шифратор-дешифратор модели с механизмом внимания. Были использованы различные архитектуры нейронных сетей, такие как: обычный перцептрон, LSTM и их модификации, остаточные сверточные сети ResNet. Полученная ошибка распознавания слов и символов оказалась немного больше, чем результаты, полученные с помощью базовых решений. Но было показано, что интегральные модели могут работать и без языковых моделей для русского языка, при этом показывая выше среднюю скорость декодирования.

В будущем планируется провести эксперименты с другими типами признаков, выполнить интеграцию языковых моделей в модели с механизмом внимания и проведение экспериментов с методом передачи данных.

Работа выполнена при финансовой поддержке фонда РФФИ (проект № 18-07-01216) и Совета по грантам Президента РФ (проект МК-1000.2017.8).

ЛИТЕРАТУРА

1. Miao Y., Gowayyed M., Metze F. EESEN: End-to-end speech recognition using deep RNN models and WFST-based decoding. *Automatic Speech Recognition and Understanding (ASRU)*, 2015 IEEE Workshop on, IEEE. 2015. pp. 167–174.
2. Popovi´c B., Pakoci E., Pekar D. End-to-End Large Vocabulary Speech Recognition for the Serbian Language. *International Conference on Speech and Computer*, Springer. 2017. pp. 343–352.
3. Mohri M., Pereira F., Riley M. Weighted finite-state transducers in speech recognition. *Computer Speech & Language.* 2002. vol. 16. no. 1. pp. 69–88
4. Allauzen C., et al.: A general and efficient weighted finite-state transducer library. *Implementation and Application of Automata.* 2007. pp. 11–23.
5. Soltau H., Liao H., Sak H. Neural speech recognizer: Acoustic-to-word LSTM model for large vocabulary speech recognition. URL: https://arxiv.org/abs/1610.09975. 2016. preprint: arXiv: 1610. 09975.
6. Liao H., McDermott E., Senior A. Large scale deep neural network acoustic modeling with semi-supervised training data for YouTube video transcription. Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on, IEEE. 2013. pp. 368–373.
7. Сборник видеоресурсов. URL: https://www.youtube.com/yt/lineups/ (дата обращения 13.07.2018).
8. Bahdanau D., et al. End-to-end attention-based large vocabulary speech recognition. Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on, IEEE. 2016. pp. 4945–4949.
9. Chan W., Jaitly N., Le Q., Vinyals O. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition // Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on, IEEE. 2016. pp. 4960–4964.
10. Graves, Ferna´ndez S., Gomez F., Schmidhuber J. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. Proceedings of the 23rd international conference on Machine learning, ACM. 2006. pp. 369–376.
11. Bridle J.S. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. Neurocomputing, Springer. 1990. pp. 227–236.
12. Graves, Jaitly N. Towards end-to-end speech recognition with recurrent neural networks. Proceedings of the 31st International Conference on Machine Learning (ICML-14). 2014. pp. 1764–1772.
13. Cho K., et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. URL: https://arxiv.org/abs/1406.1078. 2014. preprint: arXiv: 1406.1078.
14. Sutskever, Vinyals O., Le Q.V. Sequence to sequence learning with neural networks. Advances in neural information processing systems. 2014. pp. 3104–3112.
15. Kipyatkova, I., Karpov, A.: Dnn-based acoustic modeling for russian speech recognition using kaldi. In: International Conference on Speech and Computer. pp. 246–253. Springer (2016)
16. Verkhodanova, V., Ronzhin, A., Kipyatkova, I., Ivanko, D., Karpov, A., Zelezn`y,M.: Havrus corpus: high-speed recordings of audio-visual russian speech. In: International Conference on Speech and Computer. pp. 338–345. Springer (2016)
17. Kipyatkova, I., Karpov, A.: Lexicon size and language model order optimization for russian lvcsr. In:Zelezn´y, M., Habernal, I., Ronzhin, A. (eds.) Speech and Computer. pp. 219–226. Springer International Publishing, Cham (2013)
18. Chen, S.F., Goodman, J.: An empirical study of smoothing techniques for language modeling. Computer Speech & Language 13(4), 359–394 (1999)
19. Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., et al.: The kaldi speech recognition toolkit. In: IEEE 2011 workshop on automatic speech recognition and understanding. No. EPFL-CONF-192584, IEEE Signal Processing Society (2011)
20. Markovnikov, N., Kipyatkova, I., Karpov, A., Filchenkov, A.: Deep neural networks in russian speech recognition. In: Conference on Artificial Intelligence and Natural Language. pp. 54–67. Springer (2017)
21. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition.pp. 770–778 (2016)
22. Liang, M., Hu, X.: Recurrent convolutional neural network for object recognition. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3367–3375 (2015)
23. Ganchev, T., Fakotakis, N., Kokkinakis, G.: Comparative evaluation of various mfcc implementations on the speaker verification task. In: Proceedings of the SPECOM. vol. 1, pp. 191–194 (2005)
24. Ravindran, S., Demirogulu, C., Anderson, D.V.: Speech recognition using filterbank features. In: The Thrity-Seventh Asilomar Conference on Signals, SystemsComputers, 2003. vol. 2, pp. 1900–1903 Vol.2 (Nov 2003)
25. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. CoRR abs/1502.03167 (2015)
26. Popovi´c, B., Pakoci, E., Pekar, D.: End-to-end large vocabulary speech recognition for the serbian language. In: International Conference on Speech and Computer. pp. 343–352. Springer (2017)
27. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. CoRR abs/1409.0473 (2014)
28. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR abs/1412.6980 (2014)

N.M. Markovnikov, I.S. Kipyatkova (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg)

Research of the methods of constructing end-to-end Russian speech recognition systems

End-to-end speech recognition systems incorporating deep neural networks (DNNs) have achieved good results. We propose applying CTC (Connectionist Temporal Classification) models and attention-based encoder-decoder in automatic recognition of the Russian continuous speech. We used different neural network models such as Long short-term memory (LSTM), bidirectional LSTM and Residual Networks to provide experiments. We got recognition accuracy a bit worse than hybrid models but our models can work without large language model and they showed better performance in terms of average decoding speed that can be helpful in real systems.

РАЗРАБОТКА СИСТЕМЫ ГЕОМЕТРИЧЕСКИХ ПРИЗНАКОВ ДЛЯ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ ВИЗУАЛЬНОЙ РУССКОЙ РЕЧИ

А. А. Аксёнов
(*ФГБУН «Санкт-Петербургский институт информатики и автоматизации Российской академии наук», 199178, Россия, С.-Петербург, 14 линия, 39*
тел. +7 921 928 86 88, E-mail: a.aksenov95@gmail.com)

Д. А. Рюмин, **Д. В. Иванько**
(*ФГБУН «Санкт-Петербургский институт информатики и автоматизации Российской академии наук», 199178, Россия, С.-Петербург, 14 линия, 39*
тел. +7 921 362-6287, Email: denis.ivanko11@gmail.com)

Аннотация

Использование видеоинформации играет важную роль для автоматического распознавания речи. В настоящее время системы распознавания речи, основанные только на аудиоинформации, достигли определенного порога точности, и многие исследователи рассматривают решение проблемы путем использования визуальной модальности для получения лучших результатов. Несмотря на то, что речевой аудиосигнал более репрезентативен, чем видео, правильное объединение этой информации может улучшить качество и надежность всей системы распознавания, что было доказано на практике. Однако пока не достигнута договоренность между исследователями об оптимальном наборе визуальных признаков речи. Особо актуальным является повышение эффективности распознавания речи в реальных условиях, при которых аудиораспознавание не справляется. Из-за этого многие исследователи начали использовать визуальную информацию о речи в своих исследованиях. Во-первых, это позволяет создавать более надежные системы (поскольку видеоинформация не зависит от акустических шумов). Во-вторых, правильное объединение модальностей позволяет использовать преимущества обеих и, в то же время, устранить их недостатки, что дает наилучшие результаты распознавания. В этой статье фокус направлен на исследование системы распознавания визуальной русской речи и представляется метод для извлечения расширенного набора геометрических признаков. Также представлены результаты экспериментов, полученных с разработанной системой чтения речи по губам. Эксперименты проводились с использованием аудиовизуальной базы данных HAVRUS. Средняя точность распознавания визем системысоставляет 40,62%. Были протестированы основные современные методы визуального распознавания речи, с применением их к непрерывной русской речи на высокоскоростных записях (200 кадров в секунду).

Ключевые слова: Чтение по губам; автоматическое распознавание речи; декодирование визуальной речи; визуальные признаки; геометрические признаки.

Введение

На сегодняшний день, автоматическое распознавание речи является быстроразвивающейся областью в IT. Это подтверждается наличием больших количеств практических приложений, которые появляются почти каждый день. На данный момент самыми популярными из существующих являются Google «Speech API», Apple «Siri», Microsoft «Cortana», Amazon «Alexa», Yandex «Alisa», принадлежащие гигантаммировойIT-индустрии, которые уже заслужили признание миллионов пользователей. Наряду с ними также существует множество практических приложений, распространяющихся в разных сферах жизни человека: автоматическая обработка входящих звонков в телефонных call-центрах, голосовое управление для бытовой техники и автомобильных навигационных систем, в сфере социальных услуг для людей с инвалидностью, в области здравоохранения, полиции, образования и т. д.

Идея использования голосового ввода для преобразования речи в текст не только очень удобна, но и более естественна для пользователей по сравнению с клавиатурным вводом. За счет широкого использования методов машинного обучения и искусственных нейронных сетей (ИНС) в системах распознавания речи позволило вывести на новый уровеньпоказателиточности и надежности. Как правило, максимизация точности распознавания достигается в результате использования каскадов ИНС, обученных на различных комбинациях акустических признаков [1]. Но,несмотря на удовлетворительные результаты, полученные при правильном обучении системы для решения определенной задачи, точность распознавания спонтанной непрерывной речи в реальных условиях, к сожалению, все еще далека от человеческих возможностей.

В то же время практические приложенияна аудио основе имеют ряд существенных недостатков. При возникновении акустических шумов точность распознавания таких систем стремительно ухудшается. Основным подходом решения этой проблемы является использование методов предварительной обработки, направленных на снижение шума во входящем сигнале. Тем не менее, в реальных условиях существует множество различных шумов: начиная с стационарного широкополосного шума в телефонном канале и заканчиваяшумом толпы людей, наряду с реверберацией. Из этого следует, что не всегда возможно снизить шум надлежащим образом, в результате чего в настоящее время проблема автоматического распознавания речи остается открытой и нерешенной. Очевидно, что в шумных условиях точность распознавания автоматических систем вновь далека от человеческих возможностей.

С другой стороны, стоит отметить - речь человека носит бимодальный характер, и люди сами обращают внимание на движение губ собеседника во время разговора [2]. Поэтому не совсем правильно ожидать от автоматической системы распознавания речитакой же высокий результат, как и у человека, так как она получает значительно меньше информации.Из-за этого многие научные центры начали использовать визуальную информацию. Во-первых, это способствуетсозданию более надежных систем (т.к. видеоинформация не зависит от акустических шумов). Во-вторых, при правильномобъединении модальностей можно получить преимущества от обоих, а также устранить их недостатки и добиться наилучших результатов распознавания.

В этой статье внимание направлено на изучение визуальной русской речи и представлен метод для извлечения расширенного набора геометрических признаков. А также представлены результаты экспериментов, полученные с разработанной системой считывания губ.

Остальная часть этой статьи организована следующим образом. Обзор состояния исследовательской области представлен в разделе 2; в разделе 3 описана базовая методология, включая локализацию области интереса (ROI) и предлагаемые визуальные признаки на основе геометрии; в разделе 4 описанапостановка и результаты экспериментов; выводы приводятся в разделе 5.

Анализ существующих работ и выбор метода для разработки геометрических признаков

Одной из первых работ, в которой исследователи пытались систематизировать существующуюинформацию о распознавании аудиовизуальной речи, является [3]. В статье показано, что визуальная модальность естественной речи гораздо меньше информативна относительно аудио, а также информации, полученной от нее, достаточно для решения простых задач (например, распознавания изолированных слов). Описание текущего состояния можно найти в работах [4, 5], посвященных визуальному распознаванию речи. А также, работы [6, 7], посвящены распознаванию аудиовизуальной речи.

В рамках статистического подхода к распознаванию речи важным элементом является представительная база данных для обучения модели. Для английской речи доступны несколько баз данных, таких как: AVICAR [8], AVLetters [9], CUAVE [10], AV-Timit [11], IBMSR [12], PRAV Corpus [13] и т. д. Однако, ситуация с русской речью намного сложнее, так как существует очень малобаз данных визуальной русской речи. В работе использована собственная база данных непрерывной русской речи, записанную с помощью высокоскоростной камеры - HAVRUS[14].

Следующим важным шагом в построении системы считывания губ является поиск области интереса, в которойдвигаются губы, соответствуя речи. Этот шаг важен, так как правильное нахождение области оказывает значительное влияние на точность распознавания. Чтобы выделить ROI, многие исследователи полагаются на модель активного внешнего вида (AAM) [15], [16], Хааропоподобная функция, основанная на улучшенной классификации структуры [17], [18], порог цвета кожи [19] и т. д.

Несмотря на многочисленные работы, исследователи так ине смогли найти лучший набор признаков, чтобы считать его общепринятым для представления визуальной речи (например, по сравнению с известными признаками MFCC(Mel-frequencycepstralcoefficients) для акустической речи). На сегодняшний день существует несколько основных типов признаков. Наиболее часто используемые из них:

1. пиксельные [20] – необработанные пиксельные данные, используемые после преобразования изображения;

2. геометрические [21] – геометрическоеположение губ извлекается как признаки;

3. основанные на движении [22] –признаки, предназначенные для описания движения;

4. на основе модели [23] –строится модель видимых артикуляторов, а параметры компактной модели используются в качестве визуальных признаков или комбинации упомянутых выше признаков [24], [25].

Также существует несколько современных методов обучения модели. Изначально наиболее распространенные методы были основаны с использованием скрытых марковских моделей (HMM) для считывания губ для распознавания аудиовизуальной речи [26]. Однако, в настоящее время, подходы, на основе нейронных сетей разных архитектур, все больше набирают популярность [27].

В этом исследовании использовался алгоритм AAM для локализации ROI, разработаны геометрические признаки, с помощью которых извлекается информация о произнесенной речи, а также многослойная нейронная сеть для классификации.

Нахождение области интереса

Наиболее ценная информация о произносимой речи находится в области рта, следовательно, важным шагом является предварительная обработка входящих видеокадров и локализация области интереса. Для этого был использован алгоритм AAM, который реализован в библиотеке компьютерного зрения с открытым исходным кодом Dlib [28]. Основной идеей алгоритма является сопоставление статистической модели формы и внешнего вида объекта, с содержанием набора лицевых точек-ориентиров. В используемом нами детекторе лиц используется классическая гистограмма ориентированных градиентов (HOG) в сочетании с линейным классификатором, пирамидой изображения и методом скользящего окна. На рисунке 1 показан поиск точек-ориентиров на изображении с помощью этого метода. Точки на лице показывают углы рта, бровей, глаз и т.д.

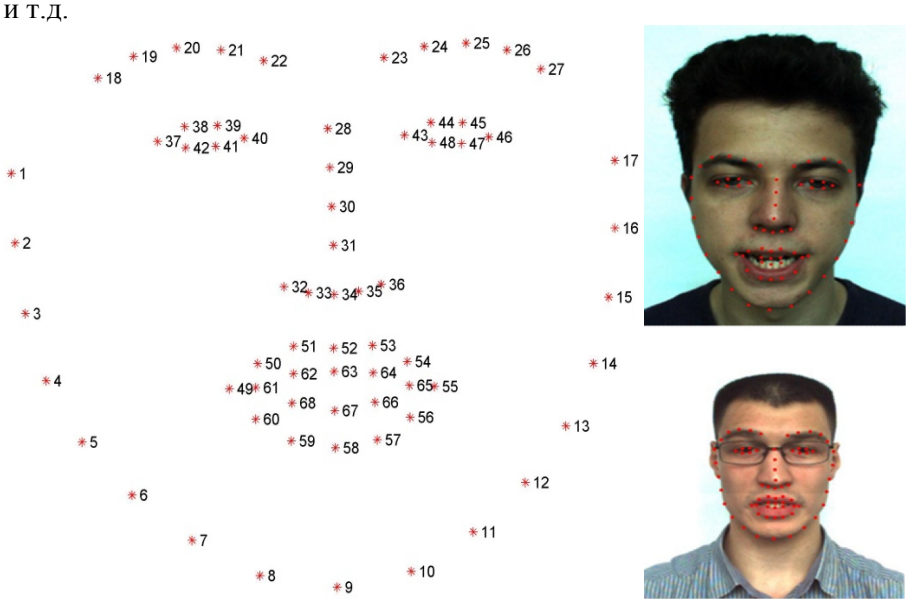


Рис. 1. Полная 2D-модель лица, используемая в [29] (слева) и результаты работы алгоритма локализации ориентиров лица (справа)

Таким образом, на каждом кадре, в котором присутствует лицо, получаются координаты лицевых точек-ориентиров, 20 из которых расположены в пределах области рта (12 на внешнихи 8 на внутренних границах губ). Этот метод отлично работает при фронтальномположении изображения лица. База данных HAVRUS содержит видеозаписив таком формате,благодаря этому, удалось получить точные координаты ориентиров губ.

Метод извлечения признаков

Общая структура метода извлечения геометрических признаков показана на рисунке 2 и включает в себя последовательное выполнение 5шагов:

1. Загрузка кадров из видеофайла.
2. Использование алгоритма обнаружения ориентиров лица, который описан в Разделе 3.1, для поиска 68 ключевых точек лица.
3. Нормализация координат, полученных ориентиров, с целью приведения данных в один формат, как это было выполнено в работе [30].
4. Вычислениеевклидовых расстояний [31] между ориентирами в соответствии с таблицей 1.
5. Сохранение вектора признаков.

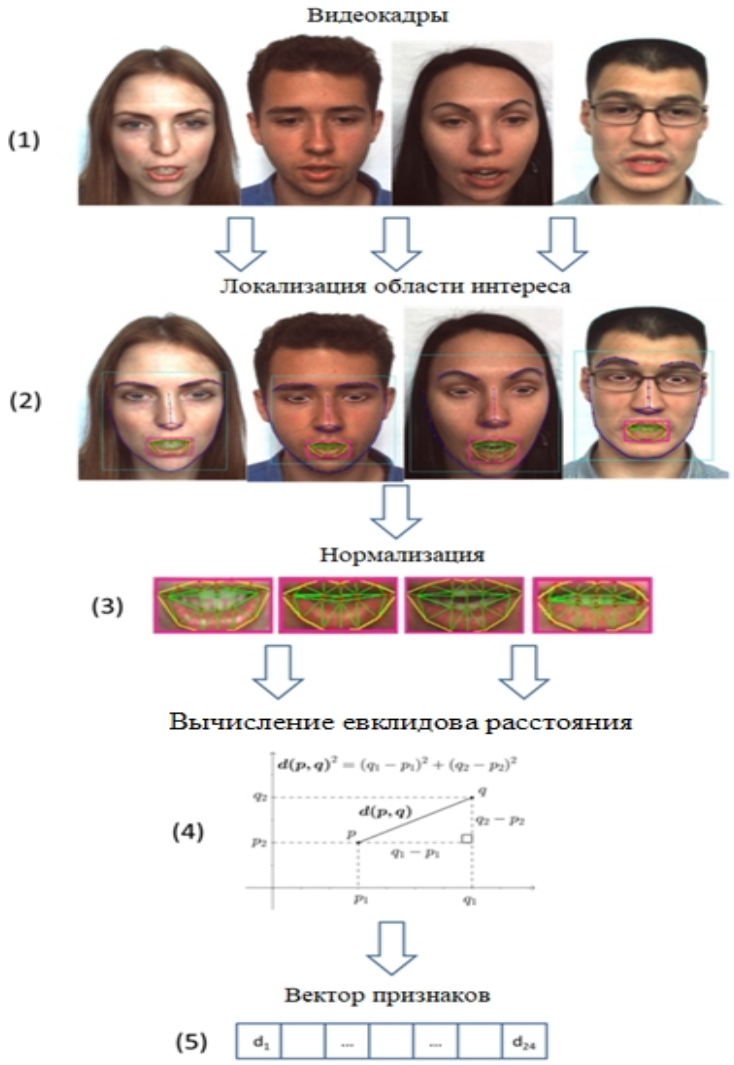


Рис. 2. Общая схема метода извлечения признаков

В этой работе определен оптимальный набор геометрических признаков, чтобы максимизировать точность распознавания. На рисунке 3 показаны 24 пары ключевых точек (выделены зеленым цветом), которые были выбраны эмпирическим путем и передают самую ценную информацию о произнесенной речи.

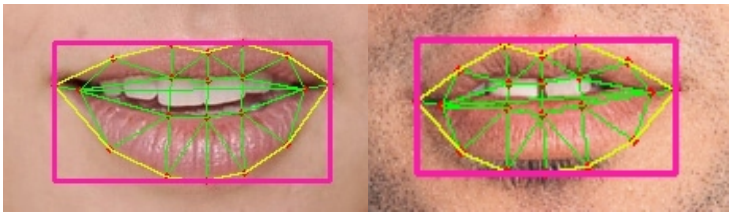


Рис. 3. Примеры обнаруженных ROI с 20 ориентирами в области рта

В таблице 1 показан набор из этих 24 выбранных признаков. В столбцахуказаны номера ключевых точек в соответствии с 2D-моделью лица (рис.1, слева), евклидово расстояние между которыми было принято в качестве признаков, например:признак # 24 показывает ширину внутренних границ губ (ориентиры 61-65). Результаты, находящиеся в экспериментальном разделе, были получены с использованием этого набора признаков.

Таблица 1

Взаимосвязь между ориентирами лица, используемая для извлечения признаков

#	Расстояниемежду ориентирами(№)	#	Расстояниемежду ориентирами (№)
1	49 - 61	13	54 - 64
2	61 - 60	14	64 - 53
3	60 - 68	15	64 - 52
4	68 - 59	16	52 - 63
5	59 - 67	17	52 - 62
6	67 - 58	18	62 - 51
7	67 - 57	19	62 - 50
8	57 - 66	20	50 - 61
9	66 - 56	21	62 - 68
10	56 - 65	22	63 - 67
11	65 - 55	23	64 - 66
12	65 - 54	24	61 - 65

Обучение многослойного перцептрона (MLP)

Для классификации визембыли взяты многослойные перцептроны (MLP), которые были обучены с помощью библиотекимашинного обучения Scikit-learn[21]. MLP -это контролируемый алгоритм обучения, который изучает функцию $f(\cdot): R^m \to R^o$ путем обучения по набору данных, где m - количество признаковнавходе, а o - количество признаковнавыходе. Учитывая набор признаков $X = x_1, x_2, \ldots, x_m$ и цель y, возможно обучениеаппроксимации нелинейной функции для классификации или регрессии. На рисунке 4 показан один скрытый слой MLP со скалярным выходом [32].

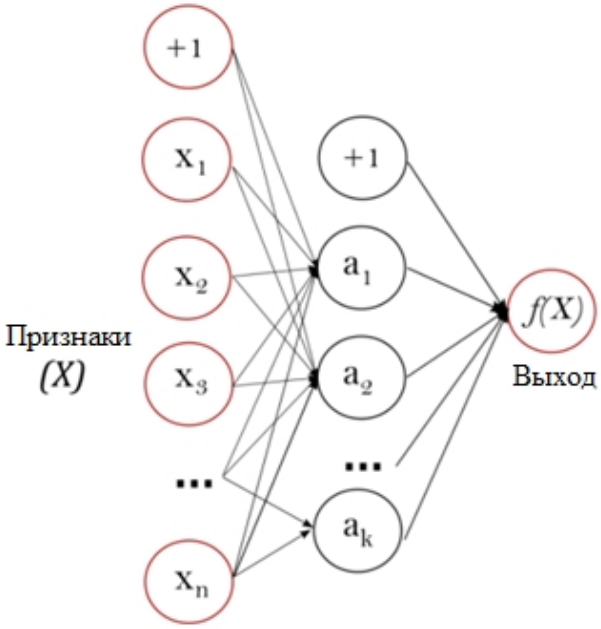


Рис. 4. MLP с одним скрытым слоем [32].

Самый левый слой, известный как входной слой, состоит из набора нейронов$x_i \ |x_1, x_2, \ldots, x_m$, представляющих входные признаков (24 в нашем случае). Каждый нейрон в скрытом слое преобразует значения из предыдущего слоя с взвешенным линейным суммированием $\omega_1 x_1 + \omega_2 x_2 + \ldots + \omega_m x_m$, за которым следует нелинейная функция активации $g(\cdot): R \rightarrow R$. Выходной уровень принимает значения из последнего скрытого уровня и преобразует их в выходные значения.

В этой работе использованы следующие основные параметры нейронной сети: в качестве функции активации использованы выпрямленная линейная единичная функция, возвращающая$f(x) = max(0, x)$. Количество нейронов в скрытом слое составляло от 1 до 100. Размер партии вычислялся по формуле $batch_{size} = min(200, n_{samples})$. Максимальное количество итераций составило 200 с допуском для оптимизации.

Результаты эксперимента

Эксперименты проводились с использованием корпуса HAVRUS, состоящего из высокоскоростных видеозаписей (200 кадров в секунду) для 20 дикторов. Каждый из них произнес 200 фраз на русском языке, взятых из фонетически насыщенных текстов. Разрешение видеоданных составляет 640 × 480 пикселей. База данных также содержит маркировку фонем и визем.

В этой работе решенатак называемая задача распознавания фонем / визем (вход системы – изображение, выход –распознанная фонема / визема). Согласно существующей маркировке HAVRUS, разделены имеющиеся данные на 48 классов в соответствии с количеством фонем в русском языке. После этого данные для каждого класса были разделены на отношение набора данных для обучения к тестовому набору данных как 75:25%. Затем былиобучены 48 многослойных перцептронов (по одному для каждого класса визем) в соответствии с разделом 3.3. Следовательно, когда на вход подается изображение из тестового набора, получается вероятность его принадлежности к определенному классу из каждого MLP и, затемвыбирается MLP с наивысшей вероятностью. Точность в этом случае означает правильное распознавание визем на тестовом наборе по сравнению с маркировкой визем в речевой базе данных.

Наилучший результат распознавания в этой системе составляет 40,62% при использовании 85 нейронов MLP. На рисунке 5 показана зависимость точности распознавания визем от количества слоев. Конечно, главная цель этой работы заключалась не в поиске оптимальной конфигурации нейронной сети. Однако наблюдается тенденция повышения точности распознавания с увеличением числа нейронов до определенного предела.

Основной задачей этого исследования являлся поиск оптимальных наборов геометрических признаков. Можно сказать, что предварительные результаты этой работы являются необходимым промежуточным шагом для улучшения существующей основы аудиовизуального распознавания речи в России [26, 33] и будут использованы для этой цели в будущих исследованиях.

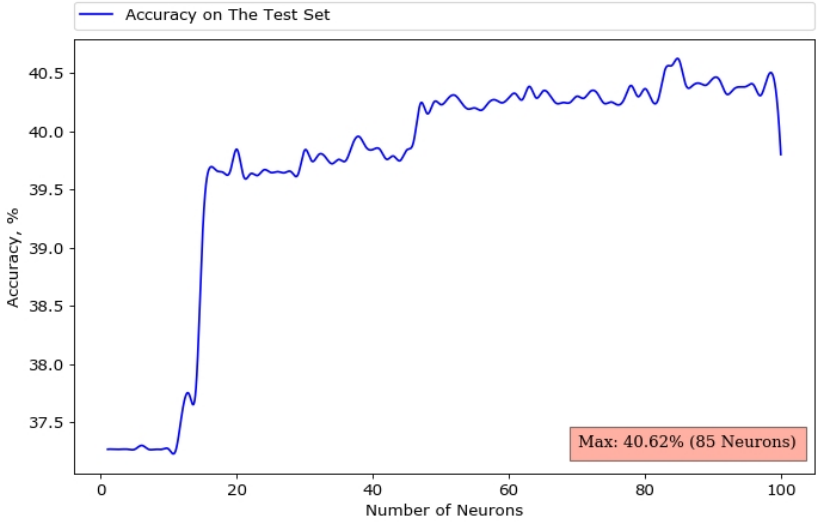


Рис. 5. Точность распознавания визем, обученных на корпусе HAVRUS

Заключение

В этой статье представлен расширенный набор геометрических признаков, предназначенных для увеличения точности системы считывания губ в русской речи, а также показаны промежуточные результаты экспериментов. Эксперименты проводились с использованием разработанной на основе MLP системы считывания губ, подготовленной библиотекой машинного обученияScikit-learn. Средняя точность распознавания системы, обученной на базе данных HAVRUS, составляет 40,62%.

Результаты этой работы будут использованы в будущих исследованиях для улучшения аудиовизуальной основы для непрерывного распознавания речи в России [26, 33]. Объединение различных визуальных признаков представляет большой практический интерес в будущих исследований.

Данное исследование выполнено при поддержке Минобрнауки России в рамках Соглашения № 14.616.21.0095 (идентификатор RFMEFI61618X0095)

ЛИТЕРАТУРА

1. Yu, D., Deng, Li.: Automatic speech recognition: a deep learning approach. Signals and Communication Technology book series, p. 315, doi: https://doi.org/10.1007/978-1-4471-5779-3 (2016)
2. McGurk, H., MacDonald, J.: Hearing lips and seeing voices. In: Nature, 264, pp. 746-748 (1976)
3. Potamianos, G., Neti, C., Matthews, I.: Audio-visual automatic speech recognition: An overview. In: Issues Audio-Visual Speech Processing, vol. 22, pp. 23 (2004)
4. Zhou, Z., Zhao, G., Hong, X., Pietikainen, M.: A Review of Recent Advances in Visual Speech Decoding. In: Image and Vision Computing, vol. 32, pp. 590-605 (2014)
5. Bowden, R., Cox, C., Harvey, R., Lan, Y., Ong, E., Owen, G., Theobald, B.: Recent developments in automated lip-reading. In: Proc. SPIE, Optics and Photonics for Counterterrorism, Crime Fighting and Defence IX, vol. 8901, pp. 13 (2013)
6. Katsaggelos, K., Bahaadini, S., Molina, R.: Audiovisual Fusion: Challenges and New Approaches. In: Proc. of the IEEE. 2015. vol. 103, no. 9, pp. 1635-1653 (2015)
7. Seong, T.W., Ibrahim, M.Z.: A Review of Audio-Visual Speech Recognition. In: Journal of Telecommunication, Electronic and Computer Engineering, vol. 10, no. 1-4, pp. 35-40 (2018)
8. Lee, B., Hasegawa, M., Goudeseune, C., Kamdar, S., Borys, S., Liu, M., Huang, T.: AVICAR: audio-visual speech corpus in a car environment. In: Proc. Interspeech 2004. pp.380-383 (2004)
9. Cox, S., Harvey, R., Lan, Y., Newmann, J., Theobald, B.: The challenge of multispeaker lip-reading. In: Proc. International conference Auditory-Visual Speech Process (AVSP), pp. 179-184 (2008)
10. Patterson, E., Gurbuz, E., Tufekci, Z., Gowdy, J.: CUAVE: a new audio-visual database for multimodal human-computer interface research. In: Proc. IEEE ICASSP 2002, vol. 2, pp.2017-2020 (2002)
11. Hazen, T., Saenko, K., La, C., Glass, J: A segment-base audio-visual speech recognizer: data collection, development, and initial experiments. In: Proc. International Conference Multimodal Interfaces, pp. 235-242 (2004)
12. Lucey, P., Potaminanos, G., Sridharan, S.: Patch-based analysis of visual speech from multiple views. In: Proc. AVSP 2008, pp. 69-74 (2008)
13. Abhishek, N., Prasanta, K.G.: PRAV: A phonetically rich audio visual corpus. In: Proc. Interspeech 2017, pp. 3747-3751 (2017)
14. Verkhodanova, V., Ronzhin, A., Kipyatkova, I., Ivanko, D., Karpov, A., Železný, M.: HAVRUS Corpus: High-Speed Recordings of Audio-Visual Russian Speech. In: Speech and Computer. SPECOM 2016. Lecture Notes in Computer Science, vol 9811, pp. 338-345 (2016)
15. Newman, J., Cox, S.: Language identification using visual features. In: Proc. IEEE Audio Speech and Language Processing, vol. 20, no. 7, pp.1936-1947 (2012)
16. Lan, Y., Theobald, B., Harvey, R.: View independent computer lip-reading. In: Proc. International Conference Multimedia Expo (ICME), pp. 432-437 (2012)
17. Zhou, Z., Hong, X., Zhao, G., Pietikainen, M.: A compact representation of visual speech data using latent variables. In: Proc. IEEE Trans. Pattern Analysis Machine Intelligent, vol. 36, no. 1, pp. 181-187 (2014)
18. Zhao, G., Barnard, M., Pietikainen, M.: Lipreading with local spatiotemporal descriptors. In: Proc. IEEE Trans. Multimedia, vol. 11, no. 7, pp. 1254-1265 (2009)
19. Estellers, V., Thiran, J.: Multi-pose lipreading and audio-visual speech recognition. In: EURALISP Journal Advanced Signal Processing, vol. 51 (2012)
20. Hong, X., Yao, H., Wan, Y., Chen, R.: A PCA Based visual DCT feature extraction method for lip-reading. In: Proc. International Conference Intelligent Information Hiding Multimedia, Signal Process, pp.321-326 (2006)
21. Cetingul, H., Yemez, Y., Erzin, E., Tekalp, A.: Discriminative analysis of lip motion features for speaker identification and speech-reading. In: Proc. IEEE Trans. Image Processing, vol. 15, no. 10, pp. 2879-2891 (2006)

22. Yoshinaga, T., Tamura, S., Iwano, K., Furui, S.: Audio-visual speech recognition using lip-movement extracted from side-face images. In: Proc. International Conference Auditory-Visual Speech Processing (AVSP). pp. 117-120 (2003)
23. Lan, Y., Theobald, B., Harvey, R., Ong, E., Bowden, R.: Improving visual features for lip-reading. In: Proc. International Conference Auditory Visual Speech Processing (AVSP). pp. 142-147 (2010)
24. Radha, N., Shahina, A., Khan, A.: An Improved Visual Speech Recognition of Isolated words using Combined Pixel and Geometric Features. In. Proc. Journal of Science and Technology, vol. 9, no. 44, pp. 7 (2016)
25. Rahmani, M.H., Alamsganj, F.: Lip-reading via a DNN-HMM hybrid system using combination of the image-based and model-based features. In: 3d International Conference on Pattern Recognition and Image Analysis. pp. 195-199 (2017)
26. Ivanko, D., Karpov, A., Ryumin, D., Kipyatkova, I., Saveliev, A., Budkov, V., Ivanko Dm., Zelezny, M.: Using a high-speed video Camera for robust audio-visual speech recognition in acoustically noisy conditions. In: SPECOM 2017, LNAI 10458, pp. 757-766 (2017)
27. Chung, J.S., Zisserman, A.: Lip reading in the wild. In: Computer Vision – ACCV 2016. pp. 87-103 (2016)
28. Implementation of computer vision library. Available at: https://github.com/davisking/dlib. (accessed 30.04.2018)
29. Baltrusaitis, T., Deravi, F., Morency, L.: 3d constrained local model for rigid and non-rigid facial tracking. In: Computer Vision and Pattern Recognition (CVPR). pp. 2610-2617 (2012)
30. Howell, D., Cox, S., Theobald, B.: Visual units and confusion modelling for automatic lip-reading. In: Image and Vision Computing, vol. 51, pp. 1-12 (2016)
31. Description of Euclidean distance calculation. Available at: https://en.wikipedia.org/wiki/Euclidean_distance (accessed 30.04.2018)
32. Machine learning toolkit. URL: http://scikit-learn.org/stable/ (accessed 30.04.2018)
33. Ivanko, D., Karpov, A., Fedotov, D., Kipyatkova, I., Ryumin, D., Ivanko, Dm., Minker, W., Zelezny, M.: Multimodal speech recognition: increasing accuracy using high-speed video data. In: Journal of Multimodal User Interfaces (JMUI) – in press.

АНАЛИТИЧЕСКИЙ ОБЗОР РЕЧЕВЫХ КОРПУСОВ ДЛЯ СИСТЕМ ОПРЕДЕЛЕНИЯ ЛОЖНЫХ РЕЧЕВЫХ СООБЩЕНИЙ

А. Н. Величко[1], **А. А. Карпов**[2], **В. Ю. Будков**[3]
(*Санкт-Петербургский институт информатики и автоматизации РАН (СПИИРАН),
С.-Петербург, 14-я линия В.О., д.39,
velichko.a.n@mail.ru, karpov@iias.spb.su, budkov@iias.spb.su*)

Аннотация

В статье представлен краткий аналитический обзор речевых и многомодальных корпусов (речевых баз данных), содержащих ложные и истинные речевые высказывания, многие из которых использовались в рамках соревнований по компьютерной лингвистике ComParE-2016. Рассматриваются следующие корпусы: Deceptive Speech Database (Университет Аризоны, США), CSC Deceptive Speech Dataset (Университет Колумбии, США), корпус Университета Ноттингема (Великобритания), корпус университета Сучжоу (Китай), многомодальный корпус Real-Life Trial Deception Detection Dataset (Университет Мичигана, США), Columbia X-Cultural Deception Corpus (Университет Колумбии, США). Рассмотренные в работе корпуса представляют собой аудио- и видеозаписи речи дикторов, являющихся носителями стандартного американского английского, британского английского, северокитайского языка (мандаринского наречия китайского языка) и русского языка. Помимо обзора баз данных также анализируются основные проблемы, с которыми сталкиваются исследователи, желающие собрать новый речевой корпус, содержащий различные паралингвистические явления.

Ключевые слова: речевой корпус, многомодальный корпус, компьютерная паралингвистика, речевые технологии, распознавание лжи в речи человека.

Введение

Важным фактором, влияющим на результат работы систем определения ложных речевых сообщений, является набор данных, на которых она была обучена. Многие ученые сталкиваются с проблемой нехватки данных, так как речевых корпусов, содержащих как ложные, так и истинные высказывания, не так много, что является естественным ввиду многих факторов. Во-первых, задача автоматического определения ложной информации в речи человека является относительно новой, хотя все больше ученых начинают проявлять интерес к исследованиям на эту тему. Во-вторых, процесс сбора аудиоданных является достаточно трудоемким и времязатратным, что, безусловно, сказывается на объемах речевых корпусов и длительности аудиозаписей. В-третьих, могут быть необходимы специально созданные сценарии, поскольку далеко не всегда является возможным провести запись речевых сообщений, содержащих ложную информацию, в реальных естественных условиях. Целью данной работы является отбор и анализ существующих на данный момент речевых и многомодальных корпусов, содержащих истинную и ложную информацию в речи человека.

Описание речевых корпусов, содержащих ложные и истинные высказывания

В последние годы возрос интерес к задаче автоматического определения ложной информации в речи, а потому в мире существует всего несколько корпусов, содержащих ложную и истинную информацию:

1) Речевой корпус Deceptive Speech Database (DSD) [1];
2) Речевой корпус CSC Deceptive Speech (CSC) [2];
3) Речевой корпус, разработанный в Университете Ноттингема [3];
4) Речевой корпус, разработанный в Университете Сучжоу [4];
5) Речевой корпус Columbia X-Cultural Deception Corpus (CXD Corpus) [5];
6) Речевой корпус, разработанный в Санкт-Петербургском институте информатики и автоматизации РАН [6].
7) Многомодальный корпус Real-Life Trial Deception Detection Dataset (RLTDDD) [7].

[1] Программист лаборатории речевых и многомодальных интерфейсов
[2] Доктор технических наук, доцент, заведующий лабораторией речевых и многомодальных интерфейсов
[3] Кандидат технических наук, старший научный сотрудник лаборатории автономных робототехнических систем

Речевой корпус DSD, был разработан в университете Аризоны (США). Данный корпус содержит аудиозаписи общей длительностью 162 минуты речи на английском языке 72 различных дикторов. В записи аудиоданных принимали участие студенты университета, они были разделены на две группы, участники которых должны быть играть роли, определенные сценарием. Роль лжецов играли участники первой группы, они «украли» вопросы и ответы для экзамена из компьютера на кафедре. Участники второй группы играли роль честных студентов, которые вернули документ в тот же кабинет. С каждым участником были проведены интервью, которые состояли из набора открытых вопросов, подразумевающих короткие ответы (в основном «да» или «нет»). Длительность отдельных высказываний информантов варьируется от 0,1 до 236 секунд. Основные параметры данного речевого корпуса представлены в Таблице 1.

Таблица 1

Параметры речевого корпуса DSD

Корпус	Обучающая часть	Настроечная часть	Тестовая часть
Кол-во дикторов (женщины, мужчины) (виновен, не виновен)	26 (11, 15) (14, 12)	23 (11, 12) (10, 13)	- - -
Кол-во аудиофайлов (виновен, не виновен) (ложные, истинные)	572 (308, 264) (182, 390)	487 (220, 267) (130, 357)	497 - -
Длительность диалога (сек) Среднее: мин.-макс.	7.6: 0.4-220	6.3: 1.1-236	6.5: 0.1-220
Длительность речевых сообщений (сек Среднее: мин.-макс.	5.1: 0.2-214	4.0: 0.1-227	3.8: 0.1-211
Кол-во фонем Среднее: мин.-макс.	32.2: 1-1480	24.5: 1-1330	23.9: 1-1268

Корпус CSC был собран исследователями Университета Колумбии (США), некоммерческого института SRI и Колорадского Университета в Боулдере в 2013 году. Он состоит из 32 часов речи 32 дикторов, которые являются носителями стандартного американского английского (16 женщин и 16 мужчин). Все участники являлись студентами или сотрудниками Университета Колумбии. Корпус был собран с целью отличить ложные речевые сообщения от истинных, используя аудиопризнаки и методы машинного обучения. Участников проинструктировали, что они участвуют в эксперименте, предназначенном для определения людей, чьи личностные качества являются схожими с характеристикой одного из ведущих бизнесменов Америки. Так, участники выполняли задания и отвечали на вопросы в шести областях знаний. После чего им сказали, что они набрали недостаточно баллов и могут попробовать еще. Интервьюер должен был определить то, как участники справились с заданиями, но ему было запрещено задавать им вопросы, не относящиеся к заданиям. Участников попросили отмечать, был их ответ истинным или ложным, нажимая на одну из двух педалей, спрятанных под столом.

Корпус, собранный в Университете Ноттингема (Великобритания) состоит из записей речи 19 мужчин, являющихся студентами или сотрудниками университета. Все они являлись носителями британского английского и не имели каких-либо речевых или слуховых нарушений. Участникам были выданы жетоны, содержащие вербальную или графическую информацию, которую они должны были скрывать в течение всего интервью. Основное интервью состояло из нейтральных вопросов и его целью было собрать контрольные данные, содержащие истинную информацию. Другие два интервью были разработаны с целью усилить напряжение участников с помощью наводящих вопросов. Первое интервью вызвало слабый эмоциональный отклик у участников, в нем задавались общие вопросы о социально приемлемом поведении и сокрытии информации. Однако, второе интервью было более провокационным и состояло из вопросов, напрямую связанных с опросом участников об их искренности. Все вопросы были заранее записаны и воспроизводились через динамики ноутбука для уверенности в том, что вопросы будут одинаковы для каждого участника. Три интервью состояли из 20 вопросов, подразумевающих короткие ответы типа «да/нет» и не были пронумерованы. Порядок интервью был одинаковым для каждого участника, речь которого сохранялась в индивидуальном аудиофайле. В целом эксперимент занял около 75 минут.

В Университете Сучжоу (Китай) речевой корпус был собран с целью получить данные, максимально близкие к реальным условиям. Так, авторы разработали игру, где участники были разделены на две группы (А и Б). Каждый участник группы А должен был рассказать историю, а участник группы Б мог задать произвольные вопросы о ней. Истории, вопросы и ответы были разными у всех дикторов. Участники группы Б не знали, являлась ли история правдой, а потому должны были принять решение самостоятельно, анализируя ответы на вопросы. В случае, если участники угадывали, они выигрывали игру, и наоборот, если не угадывали, выигрывал участник из группы А. Если история была специально придуманной ложью (блефом), рассказчик должен был приложить все усилия, чтобы не выдать себя и не проиграть. Участники группы Б могли задавать любые вопросы, касающиеся истории, чтобы заставить рассказчика нервничать и делать ошибки. Все истории, которые оказались ложью, были отобраны в качестве записей, содержащих ложную информацию. После чего была записана обычная речь участников, рассказывавших такие истории, при этом, темы рассказов могли быть любые (о себе, хобби, жизни и пр.). Данный корпус состоит из записей речи 50 участников (25 женщин, 25 мужчин в возрасте 25-35 лет). Все участники являлись носителями северокитайского языка (мандаринского наречия китайского языка).

Корпус CXD Corpus был собран в Университете Колумбии. Он содержит речевые данные с ложными и истинными высказываниями носителей стандартного американского языка и китайского мандаринского. Авторы преследовали цель разработать методы для автоматической идентификации проявлений лжи в разговорной речи. В речи участников эксперимента были изучены акустические, просодические и лексические признаки, а также их этническая принадлежность, пол и личностные качества (участники проходили пятифакторный личностный тест NEO-FFI). Авторы разработали сценарий по типу игры и разделили участников на две группы. Было проведено биографическое интервью, которое состояло из 24-х вопросов. Участники первой группы играли роль интервьюеров и им необходимо было определить, говорит ли оппонент правду, в то время, как участники второй группы были инервьюируемыми (информантами) и должны были солгать во время ответа на случайную часть вопросов. После интервью группы менялись местами. На протяжении обоих интервью участники не имели визуального контакта, Интервьюеры записывали решение относительно правдивости ответов по шкале от 1 до 5, а интервьюируемые нажимали клавиши T (правда) или F (ложь), чтобы обозначить ответы, на которые они солгали. Данный корпус состоит из 122,5 часов речи 172 пар участников.

Корпус, собранный в Санкт-Петербургском институте информатики и автоматизации РАН содержит записи речи дикторов обоих полов, в общей сумме 400 аудиофайлов объемом 120 Мбайт и общей длительностью 34 минуты. Истинность или ложность фрагментов речи определялась с помощью полиграфа. Исследованиями определения ложности и истинности информации в речи при помощи полиграфа также занимались авторы работ [8,9], для которых также были собраны речевые данные. Информация об этих данных не известна.

Существует также многомодальный корпус, RLTDDD, содержащий истинные и ложные речевые сообщения. Он был собран в Университете Мичигана (США) в 2016 году. Для сбора данных были использованы англоязычные ресурсы, предоставляющие записи судебных слушаний, на которых можно определить поведение человека при произнесении ложной и правдивой информации. На аудио- и видеозаписях были точно определены подсудимый, и свидетель; достаточно четко видны их лица на протяжении большей части записи; качество видео является достаточно хорошим, чтобы определить мимику; также, качество аудиозаписей достаточное, чтобы слышать голоса и понимать, что говорят люди. Аудио- и видеозаписи были размечены экспертом как «лживые» и «правдивые» на основании вердикта суда: виновен, не виновен или оправдан. Так, если вердикт «виновен», аудио и видеозаписи, содержащие ложную информацию, состоят из записей подсудимых, а видео, содержащие правдивую информацию – из записей свидетелей в тех же слушаниях. База данных содержит 121 запись судебных заседаний, включая 61 - с ложными данными и 60 - с истинными. Средняя продолжительность видеозаписей по всей базе данных составляет 28 сек., включая 27,7 сек. и 28,3 сек. для ложных и правдивых высказываний, соответственно. Данные содержат аудио- и видеозаписи 21 женщины и 25 мужчин, возраст информантов находится между 16 и 60 годами. Основные параметры корпуса представлены в Таблице 2.

В таблице 3 приведено сравнение вышеперечисленных речевых корпусов по следующим параметрам: модальность (одна или несколько), язык дикторов, доступ (свободный, по запросу, закрытый), количество дикторов, длительность записей.

Таблица 2

Параметры многомодального корпуса RLTDDD

Параметры	RLTDDD данные
Общее количество видеозаписей (ложные, истинные)	121 (61, 60)
Количество участников (женщины, мужчины)	46 (21, 25)
Возраст участников (лет)	16-60
Средняя продолжительность видеозаписей (сек.) (ложные, истинные)	28 (27.7, 28.8)

Таблица 3

Сравнение параметров корпусов

Корпус	Модальность	Язык дикторов	Количество дикторов	Длительность записей (мин)	Доступ
Корпус DSD	одна (аудиоданные)	английский	72	162	по запросу
Корпус CSC	одна (аудиоданные)	стандартный американский английский	32	1920	по запросу
Корпус, разработанный в Университете Ноттингема	одна (аудиоданные)	британский английский	19	нет данных	нет данных
Корпус, разработанный в Университете Сучжоу	одна (аудиоданные)	северокитайский язык (мандаринское наречие китайского языка)	50	нет данных	свободный доступ
Корпус CXD	одна (аудиоданные)	стандартный американский английский	344	7350	нет данных
RLTDDD	несколько (видеоданные)	английский	46	нет данных	свободный доступ

Заключение

Для систем автоматического определения ложной информации в речи важным фактором, влияющим на результат их работы, являются данные, на которых они были обучены. В статье были отобраны и проанализированы существующие на данный момент речевые и многомодальные корпуса, содержащие истинную и ложную информацию в речи человека. Поскольку данная тема начала набирать популярность относительно недавно, рассмотренные корпуса являются небольшими по объему. Авторами были предложены различные сценарии для того, чтобы создать наиболее реальные условия при сборе данных.

Значительная часть рассмотренных корпусов содержит речь носителей английского языка, однако они могут быть использованы в исследованиях по определению ложной или истинной информации в речи для языков из родственных языковых групп. Данная гипотеза подтверждается тем, что, хотя в различных языках и культурах существуют различия в поведении при произнесении лжи [10], языки и культуры из родственных языковых групп могут иметь схожие черты [11]. Более того, некоторые из корпусов можно найти в свободном доступе (корпус, разработанный в Университете Сучжоу и RLTDDD), однако, для доступа к большей части необходимо осуществить запрос у авторов.

Работа выполнена при поддержке Российского научного фонда (проект № 18-11-00145).

ЛИТЕРАТУРА

1. Schuller B., Steidl S., Batliner A., Hirschberg J., Burgoon J. K., Baird A., Elkins A., Zhang Y., Coutinho E., Evanini K. The INTERSPEECH 2016 Computational Paralinguistic Challenge: Deception, Sincerity & Native Language. Proceedings of INTERSPEECH-2016, San Francisco, USA, 2016, P. 2001-2005.
2. Hirschberg J., Benus S., Brenier J., Enos F., Hoffman S. F., Gilman S., Girand S., Graciarena M., Kathol A., Michaelis L. A., Pellom B. L., Shriberg E., Stolcke A. Distinguishing Deceptive from Non-Deceptive Speech. Proceedings of INTERSPEECH-2005, Lisbon, Portugal, 2005, P. 1833-1836.
3. Kirchhubel C., Stedmon A., Howard D. M. Analyzing Deceptive Speech. EPCE-2013, Springer LNCS, 2013, vol. 8019, pp. 134-141.
4. Pan X., Zhao H., Zhou Y. The Application of Fractional Mel Cepstral Coefficient in Deceptive Speech Detection. PeerJ, 2015, https://doi.org/10.7717/peerj.1194.
5. Levitan S. I., An G., Wang M., Mendels G., Hirschberg J., Levine M., Rosenberg A. Cross-Cultural Production and Detection of Deception from Speech. Proceedings of ACM Workshop on Multimodal Deception Detection, Seattle, USA, 2015, pp. 1-8.
6. Будков В.Ю., Савельев А.И., Басов О.О., Ронжин А.Л. Корпус русской речи для исследования истинности передаваемого сообщения // «Анализ разговорной русской речи 2017» труды седьмого междисциплинарного семинара, Санкт-Петербург, Санкт-Петербургский государственный университет, 2017, С. 21-25.
7. Pérez-Rosas V., Abouelenien M., Mihalcea R., Burzo M. Deception detection using re-al-life trial data. Proceedings of the 2015 ACM International Conference on Multimodal Interaction, Seattle, USA, 2015, P. 59-66.
8. Савченко В. В., Васильев Р. А. Анализ эмоционального состояния диктора по голосу на основе фонетического детектора лжи // Научные ведомости Белгородского государственного университета. 2014. Вып. 32/1. № 21(192). С. 186–195.
9. Раисов М. Э., Мещеряков Р. В. Полиграф на основе речевого ввода // Научная сессия ТУСУР-2009. Томск: В-Спектр, 2009. Ч. 3. С. 344–346.
10. Pérez-Rosas, V., Mihalcea, R. Cross-cultural Deception Detection. 52nd Annual Meeting of the Association for Computational Linguistics. In: Proceedings of ACL 2014 2, 440-445 (2014).
11. Kaya, H., Karpov, A. Efficient and effective strategies for cross-corpus acoustic emotion recognition. Neurocomputing 275, 1028-1034 (2018).

A. N. Velichko, A. A. Karpov, V. Y. Budkov (SPIIRAS, Saint-Petersburg)

Analytical survey of speech corpora for deception detection systems

This paper presents a short survey of speech and multimodal corpora (speech databases) that include speech samples of deceptive and truthful statements. Most of those corpora were used whithin the framework of the Computational Paralinguistic Challenge ComParE-2016. The paper deals with the following corpora: Deceptive Speech Database (the University of Arizona, USA), CSC Deceptive Speech Dataset (Columbia University, USA), speech corpus of the University of Nottingham (Great Britain), speech corpus of Soochow University (China), multimodal corpus Real-Life Trial Deception Detection Dataset (the University of Michigan, USA), Columbia X-Cultural Deception Corpus (Columbia University, USA). Those corpora consist of audio and video recordings of speech of native speakers (Standard American English, British English, Mandarin Chinese). Besides of the survey, we analyzed main problems faced by researches who wished to create a new corpus that contains some paralinguistic phenomena.

АНАЛИТИЧЕСКИЙ ОБЗОР ПОДХОДОВ К АВТОМАТИЧЕСКОМУ РАСПОЗНАВАНИЮ ВОЗРАСТА ДИКТОРА ПО ГОЛОСУ

М. В. Маркитантов[1], **А. А. Карпов**[2]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), Санкт-Петербург, 199178, 14-я линия В. О., 39,
тел.: +7(812)-328-04-21, m.markitantov@yandex.ru, karpov@iias.spb.su*)

Аннотация

Статья содержит аналитический обзор нескольких речевых корпусов, включая корпуса ELSDSR, Mandarin и aGender, который использовался в ходе соревнования Computational Paralinguistics Challenge (ComParE) в рамках международной конференции INTERSPEECH 2010, а также анализ используемых низкоуровневых признаков, которые содержат интонацию, интенсивность, мел-частотные кепстральные коэффициенты, а также различные наборы инструментов для извлечения признаков, такие как openSMILE и librosa. Кроме того, статья содержит анализ машинных классификаторов, таких как метод опорных векторов, глубокие нейронные сети, смесь нормальных распределений и их различные модификации. В работе исследуются как категориальные методы классификации, так и регрессионные.

Ключевые слова: компьютерная паралингвистика, машинное обучение, распознавание возраста диктора по голосу, речевые технологии.

Введение

В повседневном общении люди используют не только вербальную (речевую, текстовую и т.д.), но и невербальную (паралингвистическую, жестовую и т.д.) информацию. Последняя может содержать такие характеристики диктора, как его эмоциональное состояние, возраст, пол, наличие заболевания и другие характеристики его текущего состояния. При отсутствии прямого контакта с клиентом (пользователем), паралингвистическая информация о нем может быть полезна для предоставления определенных товаров и услуг в Интернете.

Автоматическое распознавание возраста диктора необходимо для систем верификациии идентификации дикторов, в частности, для улучшения человеко-машинного взаимодействия, а также в работе контакт-центров, учреждений здравоохранения и для повышения эффективности целевой рекламы. Автоматическая система распознавания возраста диктора может быть полезным инструментом в судебно-медицинских приложениях; она может помочь сузить список подозреваемых, когда в качестве доказательства используются речевые данные. Другие примеры коммерческих применений распознавания возраста диктора включают интеллектуальные комнаты и дома, автомобильные системы, которые могут адаптироваться к потребностям целевого пользователя.

Аналитический обзор речевых корпусов

Существует несколько корпусов речи, которые содержат представительный набор различных дикторов с указанием возраста и полав метаданных.

Корпус aGender [1] был впервые предоставлен на соревновании Computational Paralinguistics Challenge (ComParE) в рамках международной конференции INTERSPEECH 2010. Он состоит из 49 часов телефонной речи, состоящей из 795 дикторов, которые делятся на обучающий (37 часов, 770 дикторов) и тестовый (12 часов, 175 дикторов) наборы. Данные хранятся в виде звуковых файлов в формате 8кГц, 8 бит на отсчет. В корпусе используются 7 возрастных групп (классов), которые представлены в таблице 1.

Таблица 1

Возрастные классы дикторов корпуса aGender

Класс	Возрастная группа	Возраст, лет	Пол
1	Дети	7-14	М+Ж
2	Молодые люди	15-24	Ж
3	Молодые люди	15-24	М

[1] Магистрант, программист
[2] Доктор технических наук, заведующий лабораторией

4	Люди среднего возраста	25-54	Ж
5	Люди среднего возраста	25-54	М
6	Пожилые люди	55-80	Ж
7	Пожилые люди	55-80	М

В работах[2, 3] использовался корпусELSDSR (English Language Speech Database for Speaker Recognition).Он содержит голосовые сообщения 22 дикторов с возрастом от 24 до 63 лет. Для тренировочного набора было записано 154 фразы, для тестового набора – 44. Речевые файлы хранятся в формате WAV, 16 кГц, 16 бит/семпл.

В [4] использовался корпус китайской речи Mandarin, который содержит записи 4062 различных дикторов, которые сгруппированы по возрасту (16-30, 31-45, 46-60 лет), полу и акценту. Общая длительность составляет около 2000 часов. Данные хранятся в виде звуковых файлов в формате 16кГц, 16 бит на отсчет. Размер корпуса – около 215 Гб.

В [5] использовались NISTSRE2008 и SRE2010, так как содержат высказывание большого количества дикторов с метаданными, включающими возраст дикторов и пол. В данной работе были сделаны некоторые ограничения на используемые данные: использовались телефонные данные дикторов с возрастом от 20 до 70 лет.

В [6] использовался корпус голландской речи проекта N-Best. Он содержит записи прямых трансляций, новостей и интервью. В корпусе представлены 555 дикторов. В таблице 2 приведена информация о корпусе голландской речи.

Таблица 2

Метаданные корпуса голландской речи

Класс	Молодые люди		Люди среднего возраста		Пожилые люди	
Пол	М	Ж	М	Ж	М	Ж
Возраст, лет	18-35	18-35	36-45	36-45	46-81	46-81
Число дикторов	85	53	160	41	191	25

Анализ аудио признаков и методов для распознавания возраста диктора

Для извлечения информативных признаков из аудиозаписей можно использовать программный инструментарий open SMILE (https://audeering.com/technology/opensmile). Данный инструментарий позволяет получить как низкоуровневые признаки (Low Level Descriptors – LLD), включающие в себя энергозависимые (сумма акустических спектров, частоту нуль-пересечений, среднеквадратичную энергию), спектральные (мел-частота, кепстральные коэффициенты, спектральная энергия, психоакустическая четкость, гармоничность) и вокализованные признаки (частота основного тона, F0, джиттер, шиммер), так и их статистические функционалы (возрастание, убывание, минимум, максимум, среднее, стандартное отклонение, моменты, процентили и т.д.).В работах[3, 5] используются мел-частотные кепстральные коэффициенты (Mel-Frequency Cepstral Coefficients – MFCC).В работе [7] использовались коэффициенты восприятия линейного предсказания с энергией сегментов и частотой основного тона. В [5] использовались i-вектора, полученные из MFCC.

Для извлечения признаков в виде мел-спектрограммы можно использовать библиотеку librosa (https://librosa.github.io), написанную на языке Python. Так в работе [8] вместо низкоуровневых дескрипторов использовались изображения мел-спектрограмм аудиофайлов (рис. 1), которые в дальнейшем обрабатывались сверточной нейронной сетью(ConvolutionalNeuralNetworks – CNN).

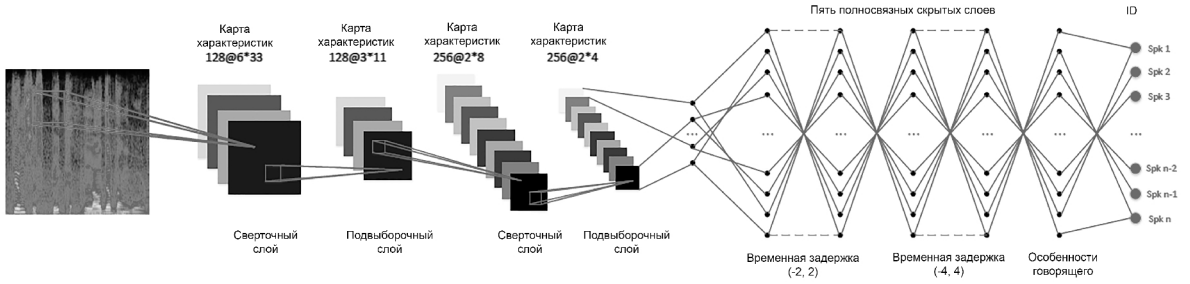


Рис. 1. Архитектура сверточной нейронной сети

Анализ достижений в области распознавания возраста диктора

В рамках международной конференции INTERSPEECH2010 в Японии, г. Макухари проводилось соревнование по направлениям компьютерной паралингвистики Computational Paralinguistics Challenge (ComPaRe)по двум темам: распознавание возраста диктора (4 класса) и пола диктора (2 класса) [9]. Участникам предоставлялись аудиозаписи корпуса a Gender и набор из 1582 признаков, полученных с помощью open SMILE.В качестве базового критерия оценки результатов применялся показатель полноты UAR (Unweighted Average Recall). Организаторы предоставили результаты испытаний базовой системы распознавания, основанной на методе опорных векторов SVM (Support Vector Machines) с линейным ядром, которая показала значения UAR для определения пола и возраста 81,21% и 48,91% соответственно, что далее использовалось в качестве минимальной планки.

Лучший результат по определению возраста составил 51,20%[7],данная система состояла из отдельных подсистем, которые обучались с использованием коротких и долгосрочных акустических и просодических признаков. Первые две подсистемы были основаны на методе SVMи нейронные сети MLP (Multi-Layer Perceptron) и использовали признаки, предоставленные организаторами. Третья подсистема также была основана на MLP с признаками, включающими коэффициенты линейного предсказания с энергий и частотой основного тона (F0). Четвертая система использовала Gaussian Mixture Models – Universal Background Model (GMM-UBM)с 1024 смесями с признаками, включающими спектры статической модуляции. Затем эти модели объединялись. Линейная логистическая регрессия и калибровка четырех систем была выполнена с помощью инструментария FoCal Multiclass Toolkit.

В работе [2] основное внимание уделяется сравнительному изучению подходов к классификации возраста диктора: классификатор Байеса, вероятностные нейронные сети, SVM, метод ближайших соседей, модели гауссовых смесей. Вероятностные нейронные сети показали наилучшие результаты. Точность классификации UARсоставила около 88,38%.

В [3] сравнивались спектральные и вокализованные признаки, классификация строилась на основе SVM. MFCCпоказали наилучший результат как по производительности, так и по точности классификации. Так же в исследовании было произведено объединение MFCCи F0, в результате точность UARсоставила 64,20%.

В [5] представлено мобильное приложение, позволяющее определять возраст диктора. В нем используется регрессионный подход. В работе[5] анализировалась нейросетевая архитектура, алгоритмы обучения, а также ансамбли нескольких сетей. Решение базировалось на i-векторах и глубоких нейронных сетях DNN (Deep Neural Network). Средние абсолютные ошибки составили 5,49 лет для женщин и 6,35 лет для мужчин.

В ряде работ используются DNN. Так, например, в [4] использовалась DNN, состоящая из четырех слоев. Была разработана система, в которой использовался детектор голосовой активности. Значение UARсоставило 85,66%.Авторы работы [10] для извлечения признаков (MFCC) и классификации используют DNN. Апостериорная вероятность softmax (многопеременная логистическая функция)в сочетании с апостериорной вероятностью последнего скрытого слоя дала точность классификации 49,50%.В [11] для извлечения признаков и классификации используется DNN.Абсолютная точность составила 57,21%. В работе [12] были предложены различные методы на основе DNN, как для параметризации аудиосигнала (BNF– Bottleneck Feature Extractor), так и классификации. Сравниваются DNNc упорядоченными и случайными весами с использованием различных признаков. В данной работе точность UARсоставила 59,95% с использованием BNF-DNN классификатора.

Заключение

В ходе исследования был проведен анализ достижений, разработок и корпусов речи в области распознавания возраста диктора по голосу. Был проведен обзор таких речевых корпусов, как ELSDSR, Mandarin и a Gender, а также анализ используемых низкоуровневыхпризнаков,которыесодержатинтонацию,интенсивность,мел-частотныекепстральныекоэффициенты, i-вектора, а также анализ наборов инструментов для извлечения признаков, таких как open SMILE и librosa, и машинных классификаторов, таких как метод опорных векторов, глубокие нейронные сети, смесь нормальных распределений и их модификации.

Различные исследования по извлечению акустических признаков и разработке машинного классификатора возраста диктора по голосу пока не позволяют получить удовлетворительную точность распознавания. Получение возрастной информации по речи диктора осложняется влиянием таких факторов, как, например, фоновый шум и голосовая вариативность. Таким образом, ключевая проблема в определении возраста диктора заключается в извлечении надежных признаков и разработке эффективного метода классификации.

Данное исследование проводится при поддержке Российского научного фонда (проект № 18-11-00145).

ЛИТЕРАТУРА

1. Burkhardt F., Eckert M., Johannsen W., Stegmann J. A Database of Age and Gender Annotated Telephone Speech// Proc. 7th International Conference on Language Resources and Evaluation (LREC 2010). Valletta. Malta. 2010.
2. Sedaaghi M.H. A Comparative Study of Gender and Age Classification in Speech Signals // Iranian Journal of Electrical and Electronic Engineering, Vol. 5, No. 1. 2009.pp. 1-12.
3. Erokyar H. Age and Gender Recognition for Speech Applications based on Support Vector Machines // Graduate Theses and Dissertations. University of South Florida. 2014.
4. Wang Z.-Q., Tashev I. Learning utterance-level representations for speech emotion and age/gender recognition using deep neural networks, // In Acoustics, Speech and Signal Processing(ICASSP). 2017 IEEE International Conference on. IEEE. 2017. pp. 5150–5154.
5. Silnova A. Exploring i-Vector Based Speaker Age Estimation // Master's Thesis. University Of Eastern Finland. 2015. P. 42.
6. Bahari M.H., Van Hamme H. Speaker age estimation and gender detection based on supervised non-negative matrix factorization // Proc. IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS). 2011. pp. 1–6.
7. Meinedo H., Trancoso I. Age and gender classification using fusion of acoustic and prosodic features // Proc. INTERSPEECH 2010.2010. pp.2818–2821.
8. Li L., Chen Y., Shi Y., Tang Z., Wang D. Deep speaker feature learning for text-independent speaker verification //Center for Speech and Language Technologies. – China. – 2017. – P. 5.
9. Schuller B., Steidl S., Batliner A., Burkhardt F., Devillers L., Mueller C., Narayanan S. The Interspeech 2010 Paralinguistic Challenge // Proc. Interspeech 2010, Makuhari, Japan, 2010.
10. Abumallouh A., Qawaqneh, Z., Barkana, B. D. Deep neural network combined posteriors for speaker's age and gender classification // CT-IETA. IEEE. 2016. pp. 1–5.
11. Qawaqneh Z., Abu Mallouh A. and Barkana B. DNN-based Models for Speaker Age and Gender Classification // Proc. 10th International Joint Conference on Biomedical Engineering Systems and Technologies, Vol.4. 2017. pp. 106-111.
12. Qawaqneh, Z., Mallouh, A. A., Barkana, B. D. Deep neural network framework and transformed MFCCs for speaker's age and gender classification //Knowledge-Based Systems, Vol.115. 2017. pp. 5–14.

M.V. Markitantov, A.A. Karpov (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, Saint-Petersburg)

An analytical review of approaches for automatical speaker's age recognition

The article contains an analytical review of several speech corpora, including ELSDSR, Mandarin and aGender, used at the Computational Paralinguistics Challenge (ComParE) in the framework ofInternational conference INTERSPEECH 2010. The article contains an analysis of low-level descriptorsused for speaker's age recognitionincluding intonation, intensity, mel-frequency cepstral coefficients and sets of software tools for extracting features, such as openSMILE and librosa. In addition, the article contains an analysis of machine classifiers, such as support vector machines, deep neural networks, Gaussian mixture models, as well as their various modifications. The paper includes an analysis of both categorical and regression classification methods.

ПОДХОДЫ К ПОСТРОЕНИЮ АССОЦИАТИВНЫХ СЛОВАРЕЙ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ АНАЛИЗА ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ТЕКСТОВ

К. В. Ненаусников[1]
(*СПИИРАН, С.-Петербург, 14 линия, 39, konstantin2113@mail.ru*)

Аннотация

Современные вопросно-ответные системы хорошо отвечают на вопросы, ответ на которые в явном виде содержится в тексте в виде утверждения, но плохо отвечают на вопросы, требующие поиска косвенной информации. Для решения этой проблемы предлагается использовать семантическое поле слов. Для этого предлагаются методы автоматического выделения ассоциативных полей из текста.

Ключевые слова: ассоциативный словарь, семантическое поле, вопросно-ответная система.

Введение

Большая часть информационных ресурсов представлена в сети Интернет, и с каждым днем ее доля все увеличивается. Поисковые системы позволяют из общего объема информации выделять страницы, содержащие данные, соответствующие требуемому запросу. Но для получения полной информации пользователю необходимо самостоятельно просматривать большинство страниц, ссылки на которые выданы в ответ на поисковый запрос, затрачивая на это свое время. Для увеличения производительности поиска информации могут использоваться вопросно-ответные системы.

Под вопросно-ответной системой понимается информационная система, способная принимать вопросы и отвечать на них на естественном языке, другими словами, это информационно-справочная система с естественно-языковым интерфейсом. На сегодняшний день существует различные типы вопросно-ответных систем. Часть из них, узкоспециализированные вопросно-ответные системы создаются для работы в ограниченной разработчиком области знания, например, медицинские вопросно-ответные системы [1]. Создание таких систем является узконаправленной задачей, требующей привлечения экспертов из рассматриваемой специальной области знания и применения специализированных алгоритмов сбора, классификации и вывода данных. Другая часть, общие (универсальные) вопросно-ответные системы, выполняют поиск ответа на вопрос пользователя по всем областям знания, наиболее известными являются системы: Exactus [2], IBM Watson [3], START [4], Яндекс Объектный ответ, Google Knowledge graph. Разработка, как общих, так и узкоспециализированных вопросных-ответных систем требует решения новых задач:

1. Построение алгоритмов быстрой обработки большого объема информации.
2. Выделение нескольких вероятных ответов и получения из них наиболее релевантного.
3. Выделение косвенной информации из текста.
4. Снятие семантической омонимии (Например, ответ на вопрос «где купить брелок египетских пирамид» не должен содержать информации о реальных пирамидах). Эта задача особенно актуальна при построении универсальных систем.

Из перечисленного списка проблем наименее разработанной является проблема выделения косвенной информации из текста[5]. Существующие вопросно-ответные системы при необходимости выделить такую информацию дают верный ответ лишь в 60 процентов случаев [6].

Одним из способов решения означенной проблемы является использование семантических словарей, например, таких как: лексическая база данных WordNet [7], толково-комбинаторный словарь И.А. Мельчука [8], семантический словарь Шведовой Н.Ю [9]. Такие словари отражают ментальные репрезентации, представление информации в голове человека (образы, пропозиции, фреймы, скрипты, сценарии и т.п.), которые могут быть выражены рассматриваемым словом с помощью специально разработанных связей между словами и возможными контекстами. Различные словари рассматривают различные семантические связи, например, синонимия, антонимия, гипонимия, гиперонимия, связь часть-целое и т.д. Такое разнообразие видов связей позволяет описывать передаваемое значение с наибольшей точностью. Выделением описанных связей занимаются эксперты. Не умаляя вклад, который оказывает создание семантических зависимостей на развитие понимания структуры передачи информации посредством

[1]Аспирант, программист.

текста, необходимо отметить, что собранные вручную словари никогда не смогут охватить весь лексический состав современного языка. Для этого необходимо разрабатывать методы автоматического выделения связей.

С точки зрения семиотики семантический смысл слова развивается, начиная с первоначального понятия, путем расширения своего семантико-смыслового содержания и потребности расширения информационного поля [10].Будем считать, что все семантические связи между словами являются ассоциативными связями. Согласно Большому психологическому словарю, ассоциативная связь – это форма связи между двумя или несколькими содержаниями сознания, выражающаяся в том, что появление одного из них влечет за собой и актуализацию другого или нескольких [11].Получение ассоциативных связей выполняется с помощью следующих экспериментов: носитель языка спонтанно реагирует на предоставленные ему словесные стимулы, затем вокруг каждого стимула составляется облако его ассоциатов (слов и словосочетаний), которые составляют ассоциативное языковое поле для рассматриваемого слова, нужно отметить, что полученное ассоциативное поле не имеет четких границ, но обладает некоторыми вероятностным распределением с однозначно определенным ядром [12]. Текст же является выражением языковой способности, поэтому связанные слова, полученные при анализе текста, условно можно приравнять к ассоциатам.

Таким образом, семантическое поле слова может быть выделено за счет анализа текстов на естественном языке, поэтому эта работа посвящена разработке методов автоматизированного построения ассоциативного словаря.

Методы автоматического выделения ассоциативных связей

Слово может выражать несколько мысленных образов. Например, слово "вода" выражает в зависимости от контекста: жидкость, пространство с водой (река, озеро и т.д.), химическое соединение и т.д. Контекст определяется близко расположенными словами и словами, глобально определяющими рассматриваемую тему, а также устойчивыми формами - фразеологизмами. Например, контекст может находиться как непосредственно в рассматриваемой фразе: "Как хочется пить, нет ли воды?", так и быть косвенно спрятанным в обсуждаемой теме: в учебнике по химии под водой скорее всего понимается химический элемент.

Задача данного исследования заключается в разработке методов выделения семантических полей слов на основе ассоциативных связей слов в тексте. Для решения этой задачи предлагается использовать смешанный подход, основанный на методах определения локальной и глобальной связанности слов.

В результате для каждого слова необходимо получить его семантическое поле, состоящее из разделенных по категориям полей ассоциаций, рисунок 1.

На рисунке 1 изображен пример для слова, содержащего три возможные ментальные репрезентации, но также существуют слова, включающие в себя большее их количество. Центральное слово содержит в себе различные ментальные репрезентации, которые могут быть выражены за счет слов ассоциаций к центральному слову. Границы ассоциативного поля размытые, поэтому одни и те же ассоциаты могут выражать различные области, но эти поля имеют вероятностный характер распределения и свое постоянное ядро.



Рис. 1. Графическое отражение структуры семантического поля слова

Предварительный этап

Для выделения семантических связей анализировалисьпроизвольные страницы сети Интернет. Предполагается, что одна страница имеет определенную тематическую окраску, поэтому каждое слово выражает единственную интерпретацию своего значения. Доступные страницы загружаются из сети Интернет с использованием технологии веб-краулера [13], выполняющего алгоритмы обхода ссылок в совокупности с headless-браузером, осуществляющим разбор исходного формата загруженного документа (PDF, HTML, MS Word и т.д.) и преобразование его в текстовый формат[14]. На данном этапе фильтруются элементы документа, не относящиеся к основной теме текста блоки информации: рекламные блоки, панели навигации и т.д.

Перед непосредственным выделением семантических полей слов, полученные тексты подвергаются графематическому анализу [15], включающему в себя определение границ абзацев, предложений и слов. Также проводится лемматизация – приведение слов к нормальной (словарной) форме и удаление стоп-слов.

Из текста выделяются наиболее часто встречаемые слова в тексте, они считаются семантическим ядром текста. Определение границы вхождения слов определяется на основании приращения рейтинга по кривой Ципфа. Экспериментально коэффициентом приращения выбрано значение 5% от максимального рейтинга в этом тексте.

Метод локальной связанности слов

В некотором тексте выбирается слово, входящее в его семантическое ядро, в дальнейшем называющееся центральным словом, для исследования его семантических связей. С целью получения локальной связанности для каждогоцентрального слова выделяются области их размещения. Полученные области анализируются, выделяя количественную оценку семантической связи, равную числу, обратному расстоянию между рассматриваемымисловами и центральным словом.Сумма связей в одном тексте считается семантическим полем.

Таблица 1 иллюстрирует работу метода на примере центральногослова «семья» в статьях Википедии «Мафия» и «Семья» [16, 17].

Таблица 1

Получение ассоциативных связей слова «семья» из статей Википедии «Мафия» и «Семья»

	Семья	**Мафия**
1	ЧЛЕН	ЧЛЕН
2	РЕБЕНОК	ДОН
3	ПСИХОЛОГИЧЕСКИЙ	ПРАВИТЬ
4	КЛИМАТ	КОНСИЛЬЕРИ
5	СЕМЕЙНЫЙ	ЧЕЛОВЕК
6	ФУНКЦИЯ	ИТАЛЬЯНЕЦ
7	ЗДОРОВЬЕ	СТРУКТУРА
8	ОПРЕДЕЛЕНИЕ	ДЖЕНОВЕЗА
9	ПОНЯТИЕ	ЧИСТОКРОВНЫЙ
10	СОЦИАЛЬНО	ОБЫЧНО
11	ТИП	ОРГАНИЗАЦИЯ
12	КАЖДЫЙ	МАФИЯ
13	НУКЛЕАРНЫЙ	БОСС
14	СОВРЕМЕННЫЙ	ИТАЛЬЯНСКИЙ
15	РОДИТЕЛЬ	ТИПИЧНЫЙ

Из примера видно, что в разных страницах присутствуют повторяющиеся элементы, которые скорее всего входят в семантическое ядро центрального слова (для уточнения нужно проанализировать большее число страниц, содержащих центральное слово «семья»), а также слова, которые переадресуют центральное слово к определенной ментальной репрезентации.

Метод глобальной связанности слов

Для получения глобальной связанности анализируется набор семантических ядер случайных текстов. Полученные списки слов объединяются: связи между одинаковыми словами складываются, для каждого слова список нормируется, процесс объединения проиллюстрирован в таблице 2.

Таблица 2

Иллюстрация объединения списка слов

Текст1	**Текст2**
Слово1	*Слово2*
Слово2	*Слово4*
Слово3	Слово6
Слово4	Слово7
Слово5	Слово8

В примере рассматриваются два текста. В каждом из столбцов представлены слова, составляющие семантическое ядро текста. В первом и во втором тексте повторяются слово 2 и слово 4, значит, количественная оценка связи между ними увеличится на условную единицу. Получается, что каждому слову соответствует список ассоциаций, встреченных в различных текстах. Этот список сортируется и нормируется.

Для получения распределённых по областям ассоциатов необходимо анализировать страницы тематических корпусов, например, национального корпуса русского языка НКРЯ [18].

Заключение

В работе предложены методы автоматического выделения ассоциативных связей на основе локальной и глобальной связанности слов. Предложенные методы позволяют определять ассоциативные связи отдельных слов, на основании чего для слова определяется семантическое поле, состоящее из нескольких ассоциативных полей, каждое из которых отражает одну из ментальных репрезентаций данного слова.

В дальнейшей работе планируется применение разработанных методов локальной и глобальной связности для автоматического создания словаря семантических полей слов, что позволит решить проблему семантической омонимии и выделения из текста косвенной информации. Это обеспечит значительный вклад в развитие опросно-ответных систем, обрабатывающих запросы на естественном языке...

Работа выполнена в рамках реализации Государственного задания на 2018 г. № 0073-2018-0002.

ЛИТЕРАТУРА

1. Миняхина П.А. Сазонова Е.С. Экспертные системы в медицине. Интеллектуальные информационные системы: тенденции, проблемы, перспективы, 2015. 105–106 С.
2. Exactus // Интеллектуальная поисковая система Exactus [Электронный ресурс]. URL: http://exactus.ru/ (дата обращения: 14.04.2018).
3. IBM Watson // IBM Watson – Build Your Cognitive Business with IBM [Электронныйресурс]. URL: http://www.ibm.com/watson/ (дата обращения: 14.04.2018).
4. START // The START Natural Language Question Answering System [Электронныйресурс]. URL: http://start.csail.mit.edu/ (дата обращения: 14.04.2018).
5. Попов С.В. Вопросно-ответная система, учитывающая семантику предметной области. ООО "Научно-издательский центр "Академия Естествознания". 2018. № 1.С. 22–26.
6. Rodrigo A., Penas A. A Study about the Future Evaluation of Question-Answering Systems. Knowledge-Based Systems. 2017. P. 3–44.
7. Miller G. A. WordNet: A Lexical Database for English. Communications of the ACM, 1995, No 11. P. 39–41.
8. МельчукИ.А.Жолковский А.К. Толково-комбинаторный словарь современного русского языка. Вена, 1984. 992 С.
9. Шведова Н.Ю. Русский семантический словарь 2007. 826 с.
10. Александров В.В. Андреева Н.А. Кулешов С.ВМетоды построения информационно-логистических систем. Санкт-Петербургский Политехнический университет. 2005. 94 с.
11. Мещеряков Б.Г. Зинченко В.П. Большой психологический словарь. Прайм-Еврознак, 2009. 816 с.
12. Караулов Ю.Н. Ассоциативный анализ: новый подход к интерпретации художественного текста. Вопросы психолингвистики. 2015. № 3(25). С. 15–35.
13. Кулешов С.В. Зайцева А.А., Марков С.В. Ассоциативно-онтологический подход к обработке текстов на естественном языке. Интеллектуальные технологии на транспорте. 2015. №4.
14. Ненаусников К.В. Кулешов С.В. Зайцева А.А. Анализ подхода создания базы данных вопросно-ответной системы на основе автоматической обработки естественно языковых текстов. Информационные технологии и телекоммуникации, 2018. Т.6 №1. 92-100 С.
15. Первушин А. модуль графематического анализа в системе обработки русскоязычных текстов // Московский институт электроники и математики НИУ ВШЕ. 2012. №15. С. 187–190.

16. Семья // Свободная энциклопедия Википедия [Электронный ресурс]. URL: https://ru.wikipedia.org/wiki/Семья (дата обращения: 10.05.2018).
17. Мафия// Свободная энциклопедия Википедия [Электронный ресурс]. URL: https://ru.wikipedia.org/wiki/Мафия (дата обращения: 10.05.2018).
18. Основной корпус // Национальный корпус русского языка [Электронный ресурс]. URL: http://www.ruscorpora.ru/search-main.html

K.V. Nenausnikov (SPIIRAS, SaintPetersburg)

The modern question-answer system responds well to questions the answer to which is contained in the text in the form of an assertion but it does not answer the question demanding the search for indirect information. To solve this problem it is suggested to use the semantic field of words. For this purpose, methods for automatically extracting associative fields from text are proposed.

АНАЛИТИЧЕСКИЙ ОБЗОР ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ ДЛЯ АВТОМАТИЧЕСКОГО ОПРЕДЕЛЕНИЯ ТОНАЛЬНОСТИ РУССКОЯЗЫЧНЫХ ТЕКСТОВ

О. В. Верхоляк[1]
(*Санкт-Петербургский институт информатики и автоматизации РАН (СПИИРАН), С.-Петербург, 14 линия В.О., 39, overkholyak@gmail.com*)

Аннотация

Автоматическое определение тональности текста, а также сентимент-анализ, является одной из актуальных задач в области обработки естественного языка, представляющая собой частный случай классификации текстов и извлечении информации. Анализ высказываний, содержащих субъективную эмоциональную оценку автора, открывает новые возможности во многих сферах человеческой жизнедеятельности. Несмотря на актуальность темы, русскоязычные ресурсы, к сожалению, отстают от западных и являются не столь развитыми. В то время, как существуют работы, посвященные системному анализу подходов к решению задачи распознавания тональности текста и методам оценки их качества, на данный момент не было обнаружено сколь-либо полного аналитического обзора существующих русскоязычных баз данных и тональных словарей. Данная работа посвящена аналитическому обзору и систематизации существующих баз данных, а также методов их сбора и аннотации, существующих в русскоязычной литературе. Были проанализированы существующие текстовые корпуса, содержащие тональную оценку высказываний, а также тональные словари, содержащие эмоционально окрашенные слова. Приведены их достоинства и недостатки, а также сравнение с зарубежными источниками.

Ключевые слова: тональность текста, тональные словари, классификация.

Введение

Автоматическое определение тональности текста, а также сентимент-анализ, является одной из актуальных задач в области обработки естественного языка, представляющая собой частный случай классификации текстов и извлечении информации. Это активно развивающееся направление компьютерной лингвистики, целью которого является определение эмоциональной окраски текста, включая положительную или отрицательную оценку автора к объекту высказывания. В качестве объектов высказывания обычно выступают продукты потребления и их атрибуты, организации и их услуги, отдельные личности, политические и экономические события, обсуждаемые в мировом сообществе и т.д.Существуют несколько типов подзадач классификации, решаемых в рамках рассматриваемого направления: двухуровневая или бинарная (положительная или отрицательная оценка), трехуровневая или тернарная (положительная, отрицательная и нейтральная), а также многоуровневая или ранжированная (например, количество звезд в отзыве).

Анализ высказываний, содержащих субъективную эмоциональную оценку автора, открывает новые возможности во многих сферах человеческой жизнедеятельности. Помимо научного интереса, определение тональности текстов представляет выгоду для бизнес-компаний в виде инструмента измерения и мониторинга удовлетворенности клиентов. Большую популярность среди политологов и журналистов приобретают методы сентимент-анализа при отслеживании общественного мнения. Все чаще системы автоматического распознавания тональности становятся востребованными в рекомендательных системах, экспертных системах, а также среди маркетологов и аналитиков.

Так как методы, применяемые в данной области, строятся на основе формальной модели языка, они являются языкозависимыми [1]. Несмотря на актуальность темы, русскоязычные информационные корпусы и ресурсы, к сожалению, отстают от западных и являются не столь развитыми [2]. Существуют инициативы, направленные на сокращение этого недостатка: например, в рамках специального направления международной конференции «Диалог» ежегодно проводятся соревнования SentiRuEval [3], целью которых является тестирование систем автоматического анализа тональности русскоязычных текстов по отношению к заданному объекту. Существуют немногочисленные работы, посвященные системному анализу подходов к решению задачи иден-

[1] Младший научный сотрудник.

тификации тональности текста [1], а также методам оценки качества анализа тональности текста [4]. Однако на данный момент не было обнаружено сколь-либо полного аналитического обзора существующих русскоязычных текстовых корпусов и тональных словарей.

Данная работа посвящена аналитическому обзору и систематизации существующих баз данных, а также методов их сбора и аннотации. Структура статьи организована следующим образом: сначала дается обзор методов сбора и аннотации баз данных тональных текстов в сети Интернет; далее следует описание существующих русскоязычных баз данных, находящихся в свободном доступе, а также других баз данных, используемых разными авторами для научных исследований; следующий раздел обобщает принципы создания тональных словарей, а также приводит несколько популярных вариантов, существующих на сегодняшний день.

Методика сбора баз данных из сети Интернет

В качестве текстовых данных для анализа тональности часто используются различные отзывы, оставленные на специализированных интернет-сервисах. При этом сбор данных вручную не представляется возможным в силу огромного количества однотипных ресурсов. Поэтому пополнение данных осуществляется автоматически с помощью специальной программы web-парсера, которая собирает информацию из иерархической HTML структуры интернет-страниц, содержащих пользовательские отзывы. Поиск происходит на множестве ссылок с одинаковой HTML структурой одного интернет-ресурса. Формально задача заключается в следующем: на данной интернет-странице P, имеющей множество текстовых объектов (например, название отзыва, объект отзыва, текст отзыва, дата отзыва и т.д.), необходимо определить отображение T объектов из P, с помощью которого осуществляется пополнение базы данных D. Это отображение должно быть применимо и для других подобных интернет-страниц P'. Подобие предполагает использование однотипной структуры данных на всем множестве страниц одного интернет-ресурса.

Сведения, собранные таким образом из Интернета, часто являются зашумленными и непригодными к использованию без предварительно обработки. Чтобы избавиться от шума, было предложено задавать диапазон длины сообщений в рамках от 10 до 800 символов, среди которых по крайней мере 50% должны быть алфавитными, и по крайней мере 30% использовать кириллицу [2]. Все URL ссылки также предлагается удалять. При обработке постов из социальных сетей не рассматриваются те, которые имеют большое количество хэштэгов или количество комментариев, меньше 2. Некоторые авторы также предлагают удалять повторные сообщения [5].

Аннотация собранных баз данных зависит от их текстового содержания. В случае, когда рассматриваются пользовательские отзывы, зачастую они сопровождаются авторскими оценками в виде звездочек от 1 до 5-10. Посты в социальных сетях также могут сопровождаться эмоциональными метками, оставленными самими авторами. В случае, когда такая разметка недоступна, применяется ручная аннотация с привлечением экспертов-лингвистов. В последнее время становится популярным прибегать к технологиям краудсорсинга в целях проведения разметки корпусов. Это позволяет избавиться от затратной работы экспертов, однако также обладает своими недостатками, о которых более подробно будет рассказано в следующих секциях.

Существующие русскоязычные базы данных

Базы данных, предоставленные на соревнованиях SentiRuEvalв 2015 и 2016 годах [3,6], являются одним из наиболее крупных собраний пользовательских отзывов в Интернете. В рамках этих соревнований были подготовлены и аннотированы отзывы и твиты о ресторанах, автомобилях, банках и мобильных операторах. В [3] длина отзывов варьировалась от коротких высказываний, состоящих из одного предложения, до длинных описаний. Корпус был вручную размечен экспертами с помощью специальной программы BRAT. Аннотация включает в себя не только полярность, но также другие аспекты высказываний – ирония, сарказм, сравнение, рекомендации и т.д. В [6] корпус состоял из постов, оставленных в социальной сети Twitter, а следовательно, длина каждого высказывания была ограничена от 1 до 140 символов. Собрание текстов получилось несбалансированным – и авторы намеренно оставили изначальное распределение положительных, отрицательных и нейтральных постов, чтобы обучать классификатор в условиях, близких к реальным. Для аннотации базы данных частично использовалась плат-

форма для краудсорсинга. Всего 112 человек из 25 городов участвовали в разметке. В разметке также участвовали 4 привлеченных со стороны эксперта и 2 организатора. Каждый твит был размечен как минимум четырьмя людьми.

Еще один крупный ресурс – русскоязычный корпус RuSentiment [2], содержащий посты общего назначения из социальной сети ВКонтакте. Данные изначально были подобраны для исследования политической предрасположенности, и содержали посты с персональных страниц пользователей, принадлежащих движениям за или против Майдана. Корпус включает в себя только те посты, которые были оставлены вне этих сообществ, а также не содержали политических слов. Он отличается от SentiRuEval тем, что все сообщения имеют различную длину, а также значительную степень зашумленности. Более того, при сборе данных не были применены техники выделения определенных тематик. Разметка корпуса проводилась шестью русскоговорящими аннотаторами.

Другая база данных содержит блог-посты популярной российской интернет-платформы LiveJournal [7]. Авторы этой базы данных проанализировали 2000 топовых блоггеров, использовав тематическое моделирование для выявления политической и социальной направленности постов. Всего в коллекцию вошло 70,710 постов. Для аннотации достаточно крупной базы данных была привлечена технология краудсорсинга. Всего 87 человек из 16 городов приняли участие в процессе разметки. Разметка проводилась на 5-ти бальной шкале от -2 (сильно негативная) и +2 (сильно позитивная) окраска. Каждый пост был разбит на абзацы, так как длинные тексты имеют тенденцию содержать различную эмоциональную окраску. Сведения об этих базах данных, количество собранных текстов, их тематика, а также количество рассматриваемых классов представлены в Таблице 1.

Таблица 1

Базы данных эмоциональных текстов в свободном доступе

	Кол-во постов	Тема	Ресурс	Кол-во классов
RuSentiment [2]	31,185	Общая	ВКонтакте	5
SentiRuEval 2015 [3]	17,628	Автомобили и рестораны		4
SentiRuEval 2016 [6]	23,595	Мобильная связь и банковское обслуживание	Twitter	3
Linis [7]	19,831	Политика и социальные вопросы	LiveJournal	5
РОМИП [8]	1,631	Кинофильмы	Imhonet.ru	2, 3, 5

В наше время достаточно просто получить доступ к открытым интернет-ресурсам, содержащим отзывы о тех или иных сервисах либо товарах. Поэтому некоторые авторы собирают находящиеся в открытом доступе отзывы самостоятельно. В таком случае в качестве разметки используются оценки, сопровождающие отзывы. Чаще всего это количество звезд в диапазоне от 1 до 5-10. Так, например, используются отзывы покупателей электроники, оставленные в сервисе «Яндекс.Маркет» [9], а также отзывы зрителей кинофильмов, оставленные в сервисах «Кинопоиск» [9] и Imhonet.net [8]. Можно встретить работы, посвященные анализу потребительских отзывов о российских банках [10]. Другие авторы применяют выборку вручную размеченных по тональности твитов [11], при этом используя инструмент сбора и анализа корпуса коротких текстов, предложенный в [12]. Есть исследования, фокусирующиеся на новостных статьях сайта Министерства Иностранных Дел Российской Федерации [13] а также новостные тексты общероссийских СМИ [14]. Различные категории сообщений, представленные в этих базах данных, включают в себя комментарии, интервью, поздравления, заявления и т.д. Сведения об этих базах данных, количество собранных текстов, их тематика, а также количество рассматриваемых классов представлены в Таблице 2.

Таблица 2

Базы данных, собранные отдельными авторами

	Кол-во постов	Тема	Ресурс	Кол-во классов
Посевкинетal. [9]	100	Электроника, кинофильмы	market.yandex.ru, kinopoisk.ru	3
Бойко [10]	1,153	Банковское обслуживание	banki.ru	2
Рубцова [11]	334,904	Общая	twitter.com	2, 3
Беляков[13]	190	Политика	mid.ru	2
Пазельская[14]	762	Новости	rbc.ru	3

Словари эмоционально окрашенных слов

Существуют несколько подходов к созданию словаря эмоционально окрашенных слов: экспертный, полуавтоматический, корпусный, метод на основе графов [5,7]. При экспертном подходе тональные словари заполняются эмоционально окрашенными словами, разбитыми на классы, вручную. Достоинство такого метода заключается в высокой точности и надежности размеченных слов, однако составленный список далеко не всегда является полным и исчерпывающим, а зачастую совсем неприменимым к анализу текстов, содержащих большое количество опечаток и внесловарных выражений. Более того, ручная разметка требует много времени и является трудоемкой. Полуавтоматические методы генерации тональных словарей используют небольшой список заранее заданных эмоциональных слов, который далее расширяют с помощью лингвистических методов, таких, как измерение семантического сходства и использование словарей синонимов/антонимов. Таким образом, полуавтоматические методы частично снимают нагрузку при создании тональных словарей. С другой стороны, существует опыт привлечения краудсорсинга в целях разметки. Однако такой подход имеет свои недостатки, в том числе низкое качество разметки ввиду низкой квалификации и мотивации участников. Разные авторы предлагают бороться с этими недостатками с помощью мотивирования аннотаторов геймификацией [15], а также фильтрацией разметки низкого качества. Существует подход, предлагающий обучение классификатора на вручную размеченных словах, имеющих эмоциональную окраску, и дальнейшее расширение словаря с помощью обученного классификатора [16]. Метод построение лексикона на основе теории графов был исследован [17]. Данный метод позволяет собирать большое количество неразмеченных данных.

RuSentiLex [18]– тональный словарь эмоциональных слов, организованный в алфавитном порядке. Лексикон был сгенерирован автоматически на основе извлечения оценочной лексики из нескольких списков, составленных в рамках лингвистико-инженерного подхода к анализу тональности [19], после чего эксперты-лингвисты анализировали употребление каждого слова, опираясь на новостные ленты как источник текстов, направленный на широкую аудиторию и передающий нормы современного русского языка. Он содержит следующие типы русскоязычных тональных слов: 1) литературные слова, явно выражающие отношение автора к объекту («замечательный»), либо передающие эмоции автора («печально»); 2) слова, не передающие субъективную оценку автора, но имеющие положительную или отрицательную общую коннотацию, например, «безработица», «терроризм», «рак» и т.д.; 3) сленговые и ругательные слова из Твиттера (за исключением нецензурной брани). Для всех слов в словаре указывается полярность (позитивная/негативная или двойная), тональность (оценка/эмоция), а также различия в тональной оценке для разных значений одного и того же слова, если таковые имеются. Авторы, однако, пришли к выводу, что большинство многозначных русских слов имеют одинаковую полярность во всех своих значениях. Если у различных значений одного и того же слова тональность отличается, для таких слов даются ссылки на понятия тезауруса РуТез (RuThes) [20].

SentiBase [21]–база данных, содержащая словосочетания тональных прилагательных с существительными, а также их пользовательские оценки. Авторы собрали 702 оценочных прилагательных, и на основе Русского Национального Корпуса составили пары-словосочетания прилагательных с существительными. Общее количество таких пар составило около 7000. Особенностью этой базы данных является то, что тональность приписывается не отдельным словам, а словосочетаниям. Авторы аргументируют это тем, что тональность высказывания в целом может отличаться от отдельных слов. Так как в общем случае тональность прилагательного будет зависеть от семантического класса употребляемого вместе существительного, каждому существительному в лексиконе приписывается значение его семантического класса. Таким образом можно проследить за изменением тональности в зависимости от семантического класса существительного, с которым оно вместе употреблено. База данных аннотирована с привлечением краудсорсинга, организованного с помощью «Краудсорсинговой игры». На момент написания данной статьи SentiBase является незавершенным проектом; процесс аннотации доступен в виде игры по ссылке. Так, например, в словаре присутствуют слова «умопомрачительный» и «сногсшибательный», но нет слова «прекрасный».

ProductSentiRus [16]– руководствуясь фактом о том, что в разных предметных областях употребляются различные наборы оценочных выражений, авторы используют совокупность статистических и лингвистических признаков для обучения модели, характеризующей предметно-

ориентированные оценочные слова. Общий словарь оценочных слов собирается из отдельно созданных словарей конкретных предметных областей. Эффективность общего лексикона для пользовательских отзывов доказана с помощью переноса модели определения тональности с одной предметной области на другую. Модель извлечения словаря создается для предметной области фильмов, а затем тестируется на других предметных областях. Для обучения модели были собраны 28773 субъективных отзыва с сайта imhonet.ru 17680 объективных описаний фильмов и 2 миллиона новостных документов, из которых вручную были отобраны и размечены 18362 слова, встречающиеся более 3 раз. При этом решалась задача бинарной классификации – разделении всех слов на два класса – оценочные и не оценочные, с помощью ансамбля алгоритмов логистической регрессии, LogitBoost и случайного леса.

Словарь эмоциональных слов, составленный Кольцовой et al. [7], состоит из комбинации ранее составленных словарей, а также слов, переведенных с английского языка и слов, добавленных экспертами как наиболее вероятно несущие эмоциональную окраску. Словарь был собран только из тех слов, которые встречались хотя бы дважды в любом из источников, указанных в списке. Из 9539 слов были отобраны 7546, относящиеся к теме политики и социологии, которые впоследствии были вручную аннотированы 87 участниками из 16 городов по шкале от -2 до +2.

Таблица 3

Словари эмоционально окрашенных слов

	Кол-во слов	Положительных слов	Отрицательных слов
RuSentiLex [18]	16,057	3,339	9,499
Linis [7]	9,539	902	1,891
ProductSentiRus [16]	4,079	-	-
TwitterLex [14]	1,734	-	-
SentiBase [21]	702	-	-

Для сравнения, в англоязычной литературе используются словари SentiWordNet [22] и SentiWords [23], которые содержат порядка 155000 слов.

Заключение

Проанализировав существующие источники русскоязычных баз данных и тональных словарей, можно сделать выводы, что действительно отечественная литература отстает от западной как в количестве доступных данных, так и в их объеме. Исследования в основном направлены на анализ полярности потребительских отзывов, однако также встречаются работы, посвященные анализу общественного мнения. Все тексты, составляющие упомянутые базы данных, были собраны в сети Интернет: на сайтах социальных сетей, новостных порталах, а также специализированных сервисах, предоставляющих возможность оставлять отзывы и комментарии. Тематика анализируемых текстов варьируется от потребительских товаров и услуг, таких, как электроника, автомобили, рестораны, кинофильмы, банковское обслуживание, до фрагментов политического и новостного содержания. Интернет-ресурсы, на которых собираются данные, включают в себя, но не ограничиваются: социальные сети ВКонтакте, LiveJournal, Twitter, специализированные сервисы сбора отзывов Маркет.Яндекс, Кинопоиск, Имхонет, banki.ru; новостные порталы МИД РФ и РБК. Чаще всего анализ тональности предполагает бинарную и тернарную классификацию, включающую классы негативной и позитивной лексики, а также нейтральные выражения. Встречаются много классовые разметки (от 5 до 10). В качестве аннотации по возможности используются оценки самих авторов (в случае работы с отзывами или постами в социальных сетях), в противном случае авторы прибегают к ручной разметке, а также краудсорсингу. При определении тональности текста зачастую используют тональные словари, которые содержат эмоционально окрашенные слова негативной и позитивной полярности, при чем как правило количество негативных слов в словаре значительно превалируют. Этот феномен объясняется тем, что в основном отзывы, оставленные в интернете, носят негативный характер. В целом, тенденция состоит в переходе к полуавтоматическим методам формирования эмоционального лексикона, а также методам на основе корпусного анализа и машинного обучения.

Данное исследование проводится при поддержке фонда РФФИ (проект № 18-07-01407) и Совета по грантам Президента РФ (грант № МД-254.2017.8).

ЛИТЕРАТУРА

1. Горбушин Д.А. et al. Системный анализ подходов к решению задачи идентификации тональности текста. Известия высших учебных заведений. Северо-Кавказский регион. Технические науки. 2016. № 2 (190). С. 36–41.
2. Rogers A. et al. RuSentiment: An Enriched Sentiment Analysis Dataset for Social Media in Russian. Computational Linguistics. 2018. P. 755–763.
3. Loukachevitch N. et al. SentiRuEval: testing object-oriented sentiment analysis systems in Russian. Dialog. 2015. Vol. 2. P. 3–13.
4. Гринченков Д.В., Горбушин Д.А. Методы оценки качества анализа тональности текста. Традиции русской инженерной школы: вчера, сегодня, завтра. 2015. С. 136–139.
5. Chetviorkin I., Loukachevitch N. Two-step model for sentiment lexicon extraction from Twitter streams. Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. 2014. P. 67–72.
6. Loukachevitch N.V., Rubtsova Y.V. SentiRuEval-2016: overcoming time gap and data sparsity in tweet sentiment analysis. Dialog. Moscow, 2016. P. 416–426.
7. Koltsova O.Y., Alexeeva S.V., Kolcov S.N. An Opinion Word Lexicon and a Training Dataset for Russian Sentiment Analysis of Social Media. Dialog. -, 2016. P. 277–287.
8. Chetviorkin I., Braslavskiy P., Loukachevich N. Sentiment analysis track at ROMIP 2011. Dialog. Bekasovo, 2012. P. 1–14.
9. Посевкин Р.В., Бессмертный И.А. Применение сентимент-анализа текстов для оценки общественного мнения. Научно-технический вестник информационных технологий, механики и оптики. 2015. Vol. 15, № 1.
10. Бойко М.В. Исследование удовлетворенности потребителей в банковской сфере на основе анализа текстовых отзывов. Вестник Уфимского государственного авиационного технического университета. 2014. Vol. 18, № 5 (66). С. 139–145.
11. Смирнова О.С., Шишков В.В. Выбор топологии нейронных сетей и их применение для классификации коротких текстов. International Journal of Open Information Technologies. 2016. Vol. 4, № 8. С. 50–54.
12. Рубцова Ю.В. Построение корпуса текстов для настройки тонового классификатора. Программные продукты и системы. 2015. Vol. 1, № 109. С. 72–78.
13. Беляков М.В. Анализ новостных сообщений сайта МИД РФ методом сентимент-анализа (статья 2). Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2016. № 4.
14. Пазельская А.Г., Соловьев А.Н. Метод определения эмоций в текстах на русском языке. Диалог. 2011. С. 17.
15. Hong Y. et al. Tower of Babel: a crowdsourcing game building sentiment lexicons for resource-scarce languages. Proceedings of the 22nd International Conference on World Wide Web. ACM, 2013. P. 549–556.
16. Chetviorkin I., Loukachevitch N. Extraction of Russian sentiment lexicon for product meta-domain. Proceedings of COLING 2012. 2012. P. 593–610.
17. Velikovich L. et al. The viability of web-derived polarity lexicons. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010. P. 777–785.
18. Loukachevitch N.V., Levchik A. Creating a General Russian Sentiment Lexicon. LREC. 2016.
19. Kuznetsova E.S., Loukachevitch N.V., Chetviorkin I.I. Testing rules for a sentiment analysis system. Proceedings of International Conference Dialog. 2013. P. 71–80.
20. Loukachevitch N., Dobrov B. RuThes linguistic ontology vs. Russian wordnets. Proceedings of the Seventh Global Wordnet Conference. 2014. P. 154–162
21. Лексикон эмоциональных прилагательных SentiBase, http:.web-corpora.net/~adaneyko/index.html
22. Baccianella S., Esuli A., Sebastiani F. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. Lrec. 2010. Vol. 10. P. 2200–2204.
23. Gatti L., Guerini M., Turchi M. SentiWords: Deriving a high precision and high coverage lexicon for sentiment analysis. IEEE Transactions on Affective Computing. 2016. Vol. 7, № 4. P. 409–421.

O.V. Verkholyak (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg)

An analytical review of textual corpora in Russian for automatic sentiment analysis

Automatic recognition of sentiment in the text, also known as tonality detection in Russian, is one of the hot topics in the area of Natural Language Processing (NLP), which reflects one of the tasks in the text classification and information retrieval. Analysis of emotionally colored texts, expressed as subjective author opinions, is a rich source of novel applications in different kinds of human daily activities. However, despite of the popularity, this research area in Russia has a gap in terms of the size and quantity of available databases. Whereas there are works devoted to the analysis and review of methods used for sentiment classification, as well as approaches to assessing its quality, to the best of our knowledge there is no comprehensive review devoted to the databases and emotional lexicon in Russian. Therefore, we devote current work to analytical review and systematization of existing databases, as well as creation and annotation procedures, focused on Russian language resources.

МЕТОДЫ ЧЕЛОВЕКО-МАШИННОГО ВЗАИМОДЕЙСТВИЯ НА ОСНОВЕ СЕНСОРНЫХ ЭКРАНОВ В СИТУАЦИОННЫХ ЦЕНТРАХ БЕЗОПАСНОСТИ

И. В. Котенко, **М. В. Коломеец**, **С. Н. Бушуев**, **А. М. Гельфанд**
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук, С.-Петербург, 14-я линия Васильевского острова, 39, ivkote@comsec.spb.ru*)

Аннотация

В статье предлагается концепция использования методов человеко-машинного взаимодействия на основе сенсорных экранов совместно со сложными моделями визуализации для мониторинга состояния защищенности в ситуационных центрах безопасности. Приводится перечень методов визуализации используемых для анализа данных в ситуационных центрах. Предлагается концепция человеко-машинного взаимодействия, которая базируется методе прямого управления.

Ключевые слова: визуальная аналитика; ситуационные центры безопасности; человеко-машинное взаимодействие; информационная безопасность; сенсорные экраны.

Введение

Одним из основных отличий ситуационных центрах безопасности является использование методов визуализации в процессах поддержки и принятия решений. Рост количества данных и типов источников информации о событиях безопасности обуславливает появление большого количества метрик, которые можно эффективно воспринимать только в графическом виде.

В тоже время возросшая производительность вычислительной техники позволяет визуализировать огромное количество данных.

Для осуществления визуального анализа используются несколько типов экранов которые условно можно разделить на три категории:

(1) небольшие экраны мобильных и носимых устройств,

(2) стандартные экраны на рабочих местах и

(3) большие экраны в виде видео-стен или проекций.

Если большие экраны используются лишь для вывода информации и анализа на верхнем уровне абстракции (обзор ситуации в целом), то небольшие и стандартные экраны используются для более глубокого визуального анализа путем масштабирования, фильтрации и детализации данных.

Применение сенсорных экранов при анализе данных меняет способы человеко-машинного взаимодействия, и как следствие, позволяет более быстро принимать решения и анализировать данные за счет дополнительных возможностей оператора. Если для мобильных и носимых устройств фреймворки человеко-машинного взаимодействия на основе сенсорных экранов существуют давно, то для экранов рабочих мест их еще предстоит сформировать.

Методы визуализации в ситуационных центрах безопасности

К настоящему моменту существует множество моделей визуализации позволяющие отобразить различные процессы безопасности. К наиболее исследованным моделям можно отнести модели на основе графов, карт деревьев и матриц [1].

Так при помощи графов атак [2] можно эффективно анализировать маршруты злоумышленника. Пример графа атак изображен на рисунке 1.

Интервальные графы (рис. 2) позволяют проанализировать временную взаимосвязь между различными событиями безопасности. В виде деревьев (рис. 3), как правило, отображают сетевую топологию.

Карты деревьев (рис. 4) позволяют эффективно анализировать метрики сети с учетом древовидной топологии. При помощи матриц (рис. 5) анализируют трафик, сканирования сети и наличие открытых портов. Матрицы на основе графов атак [3] (рис. 6) позволяют оценивать потенциальные маршруты злоумышленника с учетом уязвимостей сети.

Также появились и более специфические модели, например: совмещенные модели [4] (рис. 7), модели на основе карт Вороного [5] (рис. 8) и карт деревьев Вороного [6] (рис. 9), диаграммы Корда (рис. 10) и модели на основе гео-карт (рис. 11). Список возможных моделей можно продолжить [7].

Исследования в области новых способов человеко-машинного взаимодействия с данными моделями визуализации включают в себя как создание новых технологий (нейроинтерфейсы, сенсорные мультитач-экраны, форстач-экраны, экраны, имитирующие материю, голосовые помощники), так и разработку методов, которые позволяют более эффективно работать с моделями визуализации.

Например, в рамках исследований человеко-машинного взаимодействия был разработан принцип прямых манипуляций, наиболее простой пример реализации которого функция drag and drop которая присутствует в графических интерфейсах всех современных операционных систем.

В рамках работы с моделями на основе графов разрабатывают инструменты взаимодействия с моделью, такие как рыбий глаз или Кортезианова дисторсия.

Для бюджетных устройств виртуальной реальности актуальна разработка методик взаимодействия с виртуальной реальностью исключительно за счет отслеживания взгляда оператора.

Таким образом, способы человеко-машинного взаимодействия специфичны для моделей и технологий, которые используются для визуализации.

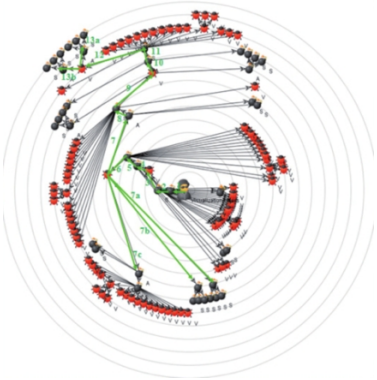


Рис. 1. Граф атак: отображает минимальные возможные пути от текущего местоположения злоумышленника (скомпрометированного хоста) до множества вероятных целей



Рис. 2. Интервальные графы: графы иллюстрирующие пересечения множества интервалов различных процессов во времени. Толщиной ребра можно задавать длительность пересечения

Рис. 3. Деревья: вид графов способный визуализировать данные иерархической структуры

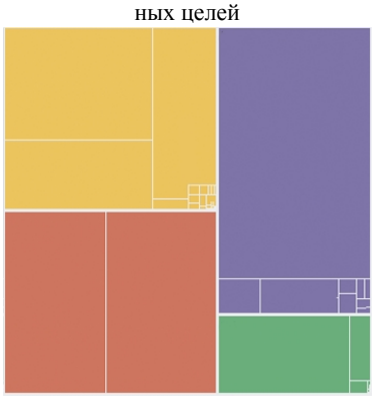


Рис. 4. Карты деревьев: множество вложенных прямоугольников, где каждый прямоугольник-предок содержит в себе множество прямоугольников-потомков. Таким образом, каждый прямоугольник на Карте Деревьев представляет собой лист дерева, а узлы деревьев представлены объединением листьев Метрики объектов можно задавать при помощи размера и иногда цвета и прозрачности прямоугольника

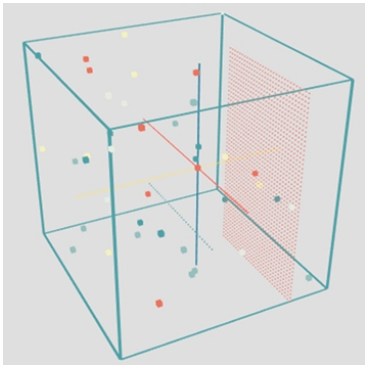


Рис. 5. Матрицы: используют пересечения сток и столбцов, которые отображают объекты. Метрики связей выражаются с использованием формы, цвета или прозрачности фигуры присутствующей на пересечении

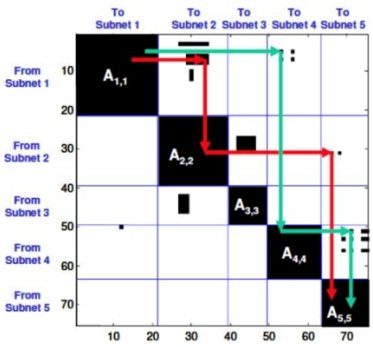


Рис. 6. Матрицы на основе графов атак: вид представления сети. Уязвимости сети, дающие возможность злоумышленнику совершить переход между хостами, выделены черным цветом. В качестве примера на рисунке изображены возможные маршруты

Стоит отметить, что сложными и многоуровневыми моделями (такими как многоуровневые графы, диаграммы Корда, карты деревьев Вороного и др.), несмотря на их явное преимущество в возможности отображения большого количества метрик, сложнее управлять. Как следствие, их эффективность не высока. Решить эту проблему могут новые методы человеко-машинного взаимодействия, которые упрощают взаимодействие и позволяют использовать сложные или нестандартные модели так же эффективно, как и простые.

В исследовательских работах, посвященных разработке методов визуализации состояния компьютерных сетей, управление моделью визуализации, как правило, происходит через данные, а не непосредственно через изображение визуальной модели. Другими словами, человеко-машинное взаимодействие происходит посредством примитивных элементов графического интерфейса пользователя: таблиц, выпадающих списков, кнопок, чекбоксов и т.д.

Такой подход разделяет данные и модель на две отличных друг от друга структуры оператор вынужден смотреть на визуальную модель, но управлять данными.

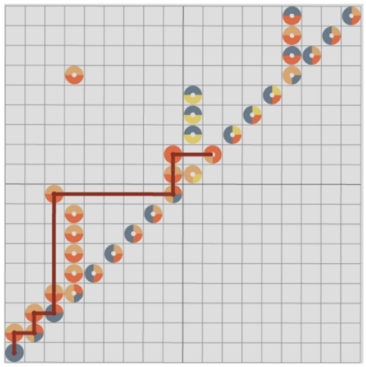


Рис. 7. Совмещенные модели: модели которые были созданы путем совмещения множества других моделей. На рисунке пример матрицы на основе графов атак где в качестве ячеек матрицы выступает модель на основе глифов



Рис. 8. Карты вороного: модель в которой объекты представлены ячейками, а связи между объектами соприкосновением ячеек. Разделение ячеек разделителем соответствует отсутствию связи. Метрики можно задавать в виде размера, цвета и прозрачности

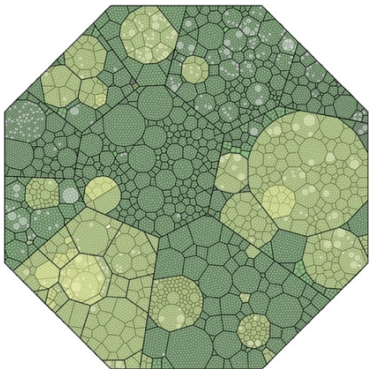


Рис. 9. Карты деревьев Вороного: состоят из вложенных друг в друга областей, которые можно задавать цветом, размером и прозрачностью. Вложенная структура соответствует иерархии, таким образом Карты Деревьев Вороного могут применяться для иерархических структур



Рис. 10. Гео-Карты: вид визуализации в которой модели накладываются на географические карты. На изображении диаграмма Вороного наложена карту Санкт-Петербурга. Расстояние от любого объекта внутри многоугольника до точки на основе которого данный многоугольник был построен всегда меньше, чем до другой точки. Данное изображение сформировано на основе координат базовых станций сотой связи сети Мегафон на Петроградском острове. Таким образом путем эксплуатации уязвимости SS7, получив CellId абонента можно узнать область в которой человек находится



Рис. 11. Диаграммы Корда - торовидная модель в которой внешний контур представляв радиальным графом, а внутренняя обычными связями

Концепция человеко-машинного взаимодействия в ситуационных центрах безопасности

В основу человеко-машинного взаимодействия лежит подход к управлению моделью визуализации посредством жестов оператора на сенсорном экране согласно принципу прямых манипуляций. Данный подход нашел большую популярность в графических интерфейсах мобильных операционных систем iOS и Android, и системах управления рабочими пространствами Mission Control (Apple MacOS) и Surface (Microsoft Windows 10).

Принцип прямых манипуляций позволяет оперировать контентом как реальными физическими объектами. Один из ярких примеров реализации данного принципа Google Material Design.

Список процессов, необходимый для комплексного анализа в рамках мониторинга, состоит из следующих процессов:

(1) формирование топологии структур анализируемых данных;
(2) анализ рисков;
(3) анализ событий и инцидентов;
(4) формирование и оценка контрмер;
(5) анализ сетевого трафика;
(6) анализ жизненного цикла элементов сети;
(7) формирование политик доступа;
(8) анализ атак.

Данные о перечисленных процессах являются многомерными структурами, которыми сложно управлять. Таким образом в рамках управления данными в ситуационных центрах необходимо использовать методы человеко-машинного взаимодействия, которые позволят работать в выделенными структурами с необходимой достаточностью и без излишней избыточности команд управления, которые будут выражаться в жестах оператора. Команды управления базируются на принципе прямых манипуляций и соответствуют структуре данных.

Приведем пример использование жестов в рамках мониторинга состояния корпоративной сети: перемещение на более низкие уровни OSI компьютерной сети в виде графа можно выполнять с помощью сведением пяти пальцев одной руки, а перемещение на более высокие уровни – разведением (рис. 12).



Рис. 12. Пример использование жестов в рамках мониторинга состояния корпоративной сети

Таким образом, модель человеко-машинного взаимодействия включает в себя список интуитивно понятных оператору жестов, которые будут определены для каждой модели визуализации для каждой структуры данных процесса обеспечения сетевой безопасности.

Сам ввод команд осуществляется посредством сенсорных мультитач-экранов. Для этого используются алгоритмы обработки жестов оператора для их трансляции в команды.

Заключение

В статье предлагается концепция использования методов человеко-машинного взаимодействия на основе сенсорных экранов совместно со сложными моделями визуализации для мониторинга состояния защищенности в ситуационных центрах безопасности.

В работе приводится перечень методов визуализации используемых для анализа данных в ситуационных центрах. Предлагается концепция человеко-машинного взаимодействия, которая базируется методе прямого управления.

Работа выполнена при поддержке РФФИ (18-07-01488) и при частичной поддержке бюджетной темы № АААА-А16-116033110102-5.

ЛИТЕРАТУРА

1. Коломеец М. В., Чечулин А. А., Котенко И. В. Обзор методологических примитивов для поэтапного построения модели визуализации данных //Труды СПИИРАН. – 2015. – Т. 5. – №. 42. – С. 232-257.
2. Igor Kotenko, Andrey Chechulin. Fast Network Attack Modeling and Security Evaluation based on Attack Graphs // Journal of Cyber Security and Mobility, Vol.3, No.1, 2014. P. 27–46.
3. Noel, Steven, et al. "Multiple coordinated views for network attack graphs." Visualization for Computer Security, 2005.(VizSEC 05). IEEE Workshop on. IEEE, 2005.
4. Kolomeec M., Gonzalez-Granadillo G., Doynikova E., Chechulin A., Kotenko I., Debar H. Choosing Models for Security Metrics Visualization // International Conference on Mathematical Methods, Models, and Architectures for Computer Network Security. – Springer, Cham, 2017. – P. 75-87.
5. Maxim Kolomeets, Andrey Chechulin, Igor Kotenko. Visualization Model for Monitoring of Computer Networks Security Based on the Analogue of Voronoi Diagrams. The International Cross Domain Conference and Workshop (CD-ARES 2016). August 31- September 2, 2016. Salzburg, Austria. F. Buccafurri et al. (Eds.): CD-ARES 2016, LNCS 9817, P.141–157, 2016.
6. Balzer, Michael, Oliver Deussen, and Claus Lewerentz. "Voronoi treemaps for the visualization of software metrics." Proceedings of the 2005 ACM symposium on Software visualization. ACM, 2005.
7. Котенко И., Новикова Е. Визуальная аналитика на страже информационной безопасности. Positive Technologies. PHD 2014. https://vimeo.com/97908798

I**.V. Kotenko, M.V. Kolomeets, S.N. Bushuev, A.M. Gelfand** (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg)

Data visualization for the identification and counteraction of undesirable information Computer-human interaction techniques based on multitouch screens in security operation centers

The article proposes the concept of using computer-human interaction methods on the basis of touch screens together with complex visualization models for monitoring the state of security in security operation centers. The paper provides a list of visualization methods used for analysis of data in security operation centers. The paper proposes a concept of computer-human interaction, which is based on the method of direct control.

АНАЛИЗ МЕТОДОВ ЧЕЛОВЕКО-МАШИННОГО ВЗАИМОДЕЙСТВИЯ В ИНСТРУМЕНТАХ ВИЗУАЛЬНОЙ АНАЛИТИКИ SIEM-СИСТЕМ

М. В. Коломеец, **И.В. Котенко**, **Н. А. Косов**, **С. А. Агеев**, А**. Ю. Иванов**
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук, С.-Петербург, 14-я линия Васильевского острова, 39, kolomeec@comsec.spb.ru*)

Аннотация

В статье рассматриваются методы человеко-машинного взаимодействия, которые применяются в визуальной аналитике в системах поддержки и принятия решений и, в частности, в SIEM-системах. В работе рассмотрены основные способы визуализации данных безопасности в SIEM-системах, а также приведены примеры использования данными системами различных методов человеко-машинного взаимодействия на примере реализации графов атак.

Ключевые слова: визуальная аналитика; SIEM-системы; человеко-машинное взаимодействие; информационная безопасность; визуализация данных.

Введение

Сложность управления безопасностью в современных инфраструктурах, характеризующихся разнородностью, ростом числа поддерживаемых платформ, приложений и клиентов, обуславливает необходимость применения автоматизированных систем мониторинга и управления безопасностью (SIEM-систем), позволяющих формировать общее видение и понимание о состоянии безопасности контролируемой системы [1-3].

Система визуализации SIEM-систем предоставляет удобный и понятный графический интерфейс для взаимодействия пользователя с функциональными компонентами, одновременно объединяя их в единую систему.

Она позволяет пользователю выполнять следующие задачи [4]:

- контролировать данные в режиме реального времени (отслеживать состояние сетевого трафика, сетевых служб, портов, доступности хостов и т.д.);
- исследовать события, хранящиеся в репозитории, для выполнения исторического анализа, формирования отчетной документации различного вида и уровня;
- создавать и редактировать правила функционирования различных модулей, например модуля корреляции событий безопасности;
- управлять инцидентами безопасности;
- управлять ресурсами сети, а также контролировать поведение пользователей SIEM-системы.

Развитие методов человеко-машинного взаимодействия позволяет повысить точность и скорость визуальной аналитики, и тем самым позволяет более быстро принимать качественные решения, которые сказываются на безопасности.

Методы визуализации

Изучение современных SIEM-систем показало, что в них наиболее часто используются двухмерные модели представления гистограммы, круговые (секторные) диаграммы и линейные графики [5-10].

С их помощью в режиме реального времени обычно отображается объем сетевого трафика, число зарегистрированных событий безопасности, наиболее "активные" хосты компьютерной сети в соответствии с рисунком 1.

Они могут быть также полезны для выявления внутренних нарушителей, показывая наиболее активных пользователей, число успешных или неудачных попыток войти в систему и т.д. в соответствии с рисунком 2.

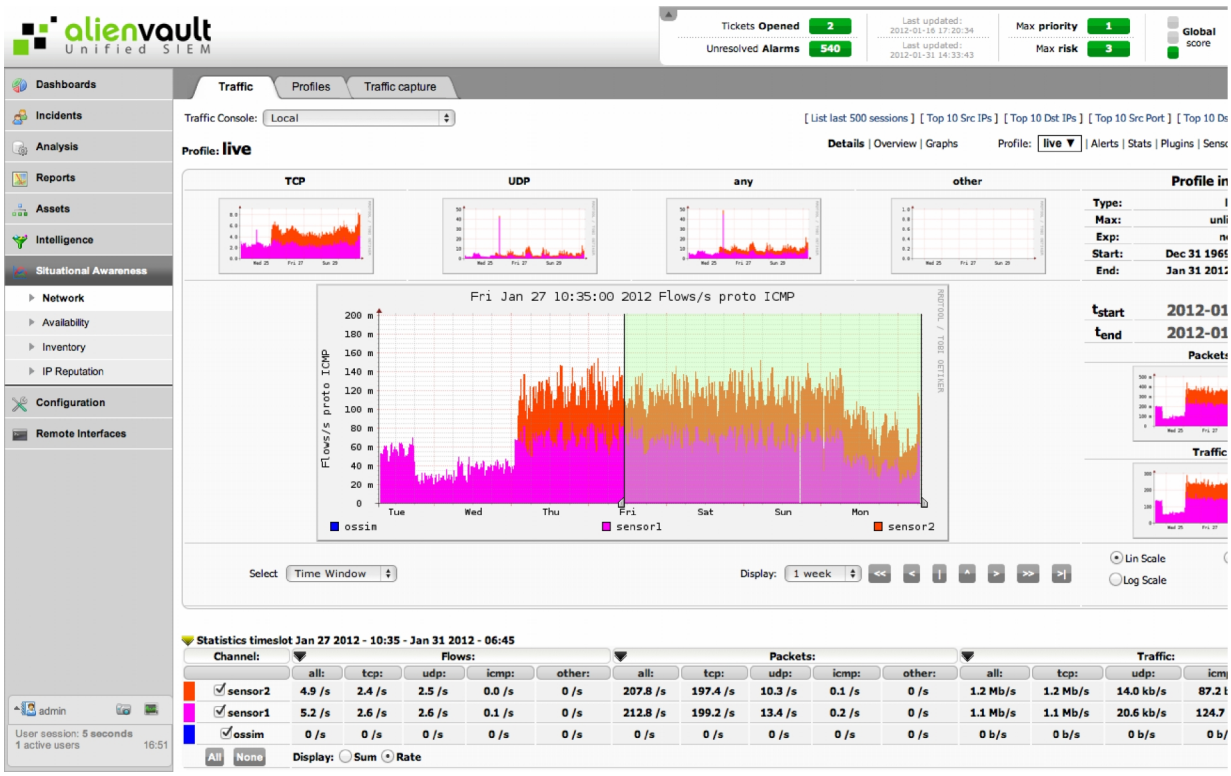


Рис. 1. Оперативная панель управления системы OSSIM [10]

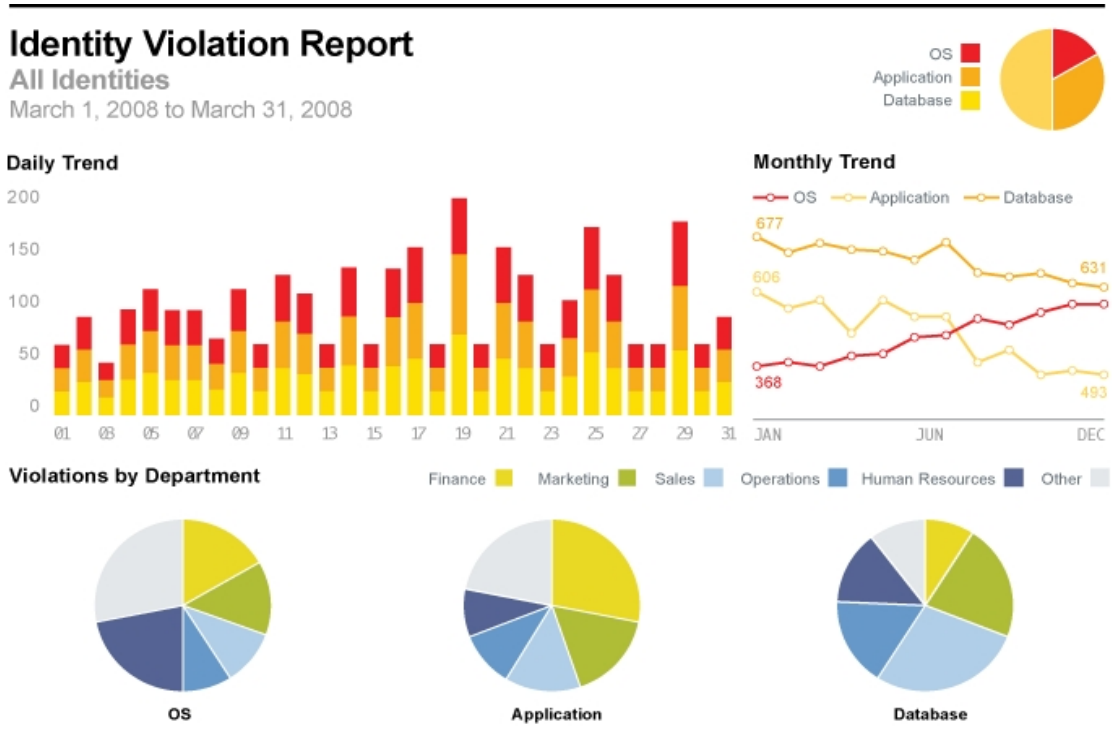


Рис. 2. Отчет о выявленных нарушениях политики безопасности, сформированный системой NetIQ Sentinel [6]

Наряду со стандартными моделями визуализации, в SIEM-системах для представления различных атрибутов сетевого трафика и метрик безопасности часто используют пиктограммы в виде цифровых приборов, которые позволяют оценить текущее значение показателя в контексте его возможных значений.

Топология компьютерных сетей и информационные потоки обычно отображаются в виде графов, вершины которых соответствуют узлам сети, а грани – связям или информационным потокам между ними [6-8].

Графы также используются для графического представления политик безопасности, правил обработки входной информации и корреляции событий безопасности, поскольку такое представление удобно для их создания, редактирования и проверки их корректности.

Однако для исследования многомерных данных большого объема необходимо применение более сложных форм графического представления данных, которые позволяют выявлять скрытые зависимости между данными, и, в конечном итоге, формулировать правила обнаружения новых информационных угроз.

Методы человеко-машинного взаимодействия

В большинстве случаев графические объекты SIEM-систем являются интерактивными, при наведении указателя мыши на элементы визуализации появляется всплывающая подсказка с детальной информацией о представленном объекте и его свойствах, кроме того, такие всплывающие подсказки часто содержат гиперссылки на исходные данные, благодаря чему упрощается процесс поиска требуемой информации.

Тем не менее, лишь некоторые SIEM-системы поддерживают три основные способа человеко-машинного взаимодействия для визуального анализа: масштабирование, фильтрацию и детализацию. Также, с внедрением новых принципов построения интерфейсов и с внедрением новых аппаратных составляющих SIEM (сенсорные экраны, носимые устройства и т.д.) появляются новые способы человеко-машинного взаимодействия, которые влияют на эффективность визуального анализа.

В результате исследования были выявлены основные способы человеко-машинного взаимодействия, которые приводятся ниже на примере реализации графов атак в SIEM-системах [11, 12].

Для взаимодействия с графическим представлением информационных объектов были определены следующие операции:

- геометрическое масштабирование;
- настройка компоновки;
- семантическое масштабирование;
- подсветка и связывание.

Геометрическое масштабирование позволяет пользователю сфокусироваться на определенных частях визуального представления информационных объектах и уменьшить их уровень связности. Основными параметрами геометрического масштабирования в графах атак являются размер области отображения (разрешение экрана) canvas.width и canvas.height, и запрашиваемые пользователем хосты geo_hosts.

Процесс масштабирования можно описать как функцию *place(canvas, ptype, host, weight)*, которая размещает в свободном месте области *canvas* объект *host*. Параметр *ptype* указывает способ компоновки графа, параметр *weight* указывает на необходимость выделить отображаемым объект (по умолчанию, *weight=normal, то есть объект не выделяется).* При этом данная функция является перегруженной и принимает в качестве объекта *host* как изображение типа хоста, так и изображение одной или нескольких из его метрик.

Настройка компоновки, то есть способа визуального представления информационных объектов, сводится к реализации различных подходов к отображению графа. Выделяют такие способы отображения, как прямолинейное (ребра графа представляются отрезками), сеточное (все вершины графа, а также точки пересечения ребер, имеют целочисленные координаты, то есть находятся в узлах координатной сетки), полигональное (ребра представляются ломанными), планарное (линии, изображающие ребра, не пересекаются между собой) и другие.

Семантическое масштабирование используется для уменьшения сложности визуального представления на основе сематических и структурных свойств информационных объектов. Например, уменьшить сложность визуального представления возможно при помощи агрегации узлов, входящих в одну подсеть или домен.

Подсветка и связывание необходимы для выделения фрагмента или пути в визуальном

представлении. Данная операция может выделять запрашиваемые хосты selected_hosts и связи selected_conn утолщенными линиями.

В работе авторов данные подходы используются для визуализации информации в разрабатываемом прототипе SIEM-системы [1, 2].

Заключение

В статье рассмотрены методы человеко-машинного взаимодействия, которые применяются в визуальной аналитике в системах поддержки и принятия решений и, в частности, в SIEM-системах. Предложены основные способы визуализации данных безопасности в SIEM-системах, а также приведены примеры использования данными системами различных методов человеко-машинного взаимодействия на примере реализации графов атак.

Работа выполнена при поддержке РФФИ (16-29-09482, 18-07-01488).

ЛИТЕРАТУРА

1. Котенко И.В., Саенко И.Б. Архитектура системы интеллектуальных сервисов защиты информации в критически важных инфраструктурах // Труды СПИИРАН. № 1 (24). 2013. С.21-40.
2. Котенко Д.И., Котенко И.В., Саенко И.Б. Методы и средства моделирования распределенных атак в больших компьютерных сетях: состояние проблемы // Труды СПИИРАН. № 3 (22). 2012. С.5-30.
3. Kotenko I., Konovalov A., Shorov A. Agent-based simulation of cooperative defence against botnets // Concurrency Computation Practice and Experience. 2012. Vol.24. № 6. P.573-588.
4. Проноза А.А., Чечулин А.А., Котенко И.В. Математические модели визуализации в SIEM-системах // Труды СПИИРАН. 2016. Вып. 46. С.90-107.
5. Arcsight SIEM [Электронный ресурс]. – 2018. – Доступно по ссылке: http://www.ndm.net/siem/main/arcsight-siem.
6. NetIQ Sentinel SIEM website [Электронный ресурс]. – 2018. – Доступно по ссылке: https://www.netiq.com/products/sentinel/.
7. QRadar SIEM website [Электронный ресурс]. – 2018. – Доступно по ссылке: http://q1labs.com/products/qradar-siem.aspx.
8. Tivoli Security Information and Event Manager websight Электронный ресурс]. – 2018. – Доступно по ссылке: http://www-01.ibm.com/software/tivoli/products/security-info-event-mgr/index.html.
9. Prelude SIEM website [Электронный ресурс]. – 2018 – Доступно по ссылке: http://www.prelude-ids.com/en/products/universal-siem/index.html.
10. OSSIM SIEM website [Электронный ресурс]. – 2018. – Доступно по ссылке: http://communities.alienvault.com/.
11. Kotenko I., Stepashkin M. Network Security Evaluation based on Simulation of Malefactor's Behavior // SECRYPT 2006 - International Conference on Security and Cryptography, Proceedings International Conference on Security and Cryptography, SECRYPT 2006. sponsors: IBM, Polytechnic Institute of Setubal. Setubal, 2006. P.339-344.
12. Котенко И. В., Новикова Е. С. Визуальный анализ защищенности компьютерных сетей //Информационно-управляющие системы. – 2013. – №. 3 (64).

M.V. Kolomeets, I.V. Kotenko, N. A. Kossov, S. A. Ageev, A. Y. Ivanov (St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg)

Analysis of human-machine interaction methods in siem-systems visual analytics instruments

The paper considers techniques of computer-human interaction that are used in visual analytics in support and decision-making systems, and in SIEM-systems. The work examines the main ways to visualize security data in SIEM-systems, as well as examples of the use of methods of computer-human interaction by these systems on the example of implementation in attack graphs.

АНАЛИТИЧЕСКИЙ ОБЗОР РОБОТОТЕХНИЧЕСКИХ СРЕДСТВ С МНОГОМОДАЛЬНЫМ ЧЕЛОВЕКО-МАШИННЫМ ИНТЕРФЕЙСОМ

И. А. Кагиров[1]
(*Санкт-Петербургский институт информатики и автоматизации Российской академии наук,
С.-Петербург, 199178, 14-я линия В.О., дом 39,
тел.: +7-(812)-328-04-21, kagirov@iias.spb.su*)

Аннотация

Статья представляет собой аналитический обзор роботизированных платформ, оснащенных многомодальным интерфейсом.Дан анализ основных тенденций применения различных каналов обмена информацией с пользователем в зависимости от сферы применения робота. Основной тенденцией оказывается минимизация усилий со стороны пользователя. Предложен проект роботизированной тележки для супермаркета, поддерживающей обмен информацией с пользователем по речевому каналу, а также при помощи русского жестового языка.

Ключевые слова: многомодальные интерфейсы, робототехника, ассистивные технологии, социальные роботы, жестовый язык.

Введение

Предметом настоящей статьи является аналитический обзор существующих на сегодня роботизированных средств, оснащенных многомодальным человеко-машинным интерфейсом. Под многомодальным интерфейсом понимается такой интерфейс, в котором скомбинированы несколько входных и выходных модальностей, например, аудио- и визуальный каналы. В данном обзоре рассматриваются роботизированные платформы, взаимодействие пользователя с которыми осуществляется несколькими способами (комбинированно или по выбору пользователя): жестами, речью, мимикой, посредством текстового ввода, сенсорно и т.д. При этом особенный интерес представляют устройства, обладающие речевым каналом обмена информацией, поскольку в повседневной жизни речевое взаимодействие чаще всего оказывается наиболее естественным и ненавязчивым для пользователя.

Все примеры роботизированных средств, оснащенных многомодальным интерфейсом[2], сгруппированы по сфере их применения, и анализируются следующие параметры: а) типы модальностей; б) сфера применения; в) роль при взаимодействии с человеком; г) внешний вид.

По данным Международной федерации робототехники [1], в мире на 2016 год насчитывалось свыше 1,8 млн. индустриальных роботов в эксплуатации. Сервисных роботов, по данным консалтинговой компании BergInsight [2], насчитывается гораздо больше, порядка 30 млн. единиц, однако бóльшая часть рынка сервисных роботов представлена роботами-пылесосами для очистки пола (порядка 24 млн. единиц), роботизированными газонокосилками (1,6 млн. единиц) и беспилотными летательными аппаратами (4 млн.). Остальные сегменты рынка сервисных роботов совокупно оцениваются в 50 тыс. единиц, сюда включены роботы-помощники, роботы-аватары, силовые экзоскелеты, хирургические роботы, автономные мобильные роботы и др. По прогнозам, предложенным в [3], в 2018-2020 гг. будет продано еще порядка 400 тыс. сервисных роботов.

На сегодня примеры успешного применения роботов можно обнаружить в разнообразных областях человеческой деятельности начиная от космических полетов и заканчивая образованием. При этом важнейшим, если не решающим фактором внедрения роботизированных устройств является естественность взаимодействия с пользователем. [4], [5].В идеале, взаимодействие человека с машиной не должно отличаться от взаимодействия двух пользователей в терминах модальности. «Главным отличием многомодальных интерфейсов от профессиональных средств взаимодействия, использующихся оператором автоматизированного рабочего места, является применение способов коммуникации, характерных для общения между людьми» [6: стр. 196]. Такой способ обмена информацией оказывается еще и достаточно удобным, так как оставляет за пользователем выбор конкретной модальности в зависимости от потребностей и коммуникативного контекста.

[1]Младший научный сотрудник.
2 См. также [7].

Разработки, освещенные в настоящем аналитическом обзоре, включают в себя не только многомодальные системы. Как было отмечено выше, одной из главных черт многомодального интерфейса – помимо наличия нескольких каналов обмена информацией –является естественность для пользователя. Поэтому в обзоревстречаютсяпримеры и такихинтерфейсов, в которых предусмотрена всего одна входная / выходная модальность, если она связана с аудио- или визуальным каналом общения, тем более что с технической точки зрения подобные системы довольно легко расширяются за счет внедрения элементов графического пользовательского интерфейса (сенсорные панели).

Ассистивные роботы

Под ассистивными роботами понимаются роботы, помогающие поддерживать на существующем уровне или повышать возможности людей, которые в силу каких-либо причин (старость, болезнь, травма, или же ограничения, вызванные рабочей средой) испытывают функциональные ограничения[8].

Чаще всего такие роботы должны быть мобильными и иметь доступ к беспроводным компьютерным сетям, что позволяет использовать их как роботов телеприсутствия и обеспечивает постоянную связь с другими людьми.

Примеров реализации ассистивных технологий в сфере сервисной робототехники довольно много. Так, компания Toyota создала робота HSR(Human Support Robot) в рамках проекта Partner Robot Family[9]. Это робот, предназначенный, в первую очередь, для облегчения выполнения некоторых бытовых задач людям, для которых перемещение в пространстве и манипуляции с окружающими их объектами доставляют серьезные трудности. Модель HSR способна поднимать и перемещать предметы весом до 1,2 кг и диаметром до 130 мм, распознавать объекты, перемещаться по помещению и ориентироваться в пространстве. Взаимодействие с пользователем осуществляется посредством графического интерфейса (в том числе и с мобильных устройств, таких как смартфоны или планшеты) или голосовыми командами. Кроме того, робот снабжен функциями телеприсутствия и удаленного контроля, что облегчает стороннюю заботу о человеке с ограниченными возможностями здоровья (ОВЗ).

Робот-компаньон Hector[10], созданный совместными усилиями ряда европейских организаций в рамках проекта Companion Able, также относится к категории ассистивных роботов. Он предназначен, в первую очередь, для одиноких пожилых людей и способен выполнять ряд бытовых задач: регулирование системы отопления, открытие и закрытие штор, включение света, напоминание о важных делах или о невыключенных бытовых приборах. Благодаря интуитивно понятному интерфейсу Hector оказывается посредником между представителями старшего поколения и современным миром, важной частью которого являются информационные технологии, не всегда доступные пожилым людям в силу их привычек.Hector способен понимать голосовые команды, откликаться на них и реагировать, перемещаясь по квартире и выполняя возложенные на него задачи, приносить небольшие предметы и служить личным органайзером. Помимо голосового управления, доступен графический интерфейс при помощи сенсорного экрана.

Еще одним ассистивным роботом является KOMPAÏ-2(2016)[11]–робот с многомодальным интерфейсом, созданный, в первую очередь, для групп населения, которым требуется сторонняя забота: одиноких лиц пожилого возраста, инвалидам и людям, страдающим различными заболеваниями, ограничивающими их физические и ментальные возможности.KOMPAÏ-2 оснащен сенсором Kinect2.0, способен распознавать лица и мимику. 3D камера и лазерный датчик позволяют ему передвигаться по жилищу, объезжать препятствия и определять своего владельца. Кроме того, робот способен отслеживать состояние подопечного, пользуясь доступом к медицинским датчикам через облачное хранилище, и вызывать доктора в случае необходимости. Устройство оснащено сенсорным монитором, что позволяет использовать робота для удаленного общения с другими лицами и обеспечивать телеприсутствие. Кроме графического интерфейса, KOMPAÏ-2 обладает устройством распознавания речи и способен выполнять голосовые команды пользователя. Наконец, на корпусе робота имеются контейнеры для медикаментов, личных вещей, поручни и тревожная кнопка, доступная даже лежащему на полу пользователю.

В рамках проекта ALIAS (AdaptableAmbientLivingAssistant [12],выполнявшегося в 2010-13 гг., была разработана робототехническая платформа, облегчающая пользователю общение со окружающим пространством. Платформа объединяет распознавание и синтез речи, нейрокомпьютерный интерфейс, распознавание лиц, а также устройства, обеспечивающие доступ в Интернет и общение посредством Skype. Робот используется как мобильная станция для общения.

Довольно интересен робот THERAPIST для нейрореабилитации пациентов, разработанный рядом испанских научно-исследовательских лабораторий [13], и способный к распознаванию и синтезу речи, мимики, жестикуляции. В настоящее время робот обладает двумя манипуляторами и мультисенсорной гуманоидной головой Muecas, имитирующей человеческую мимику. Кроме того, THERAPIST снабжен сенсором Kinect, позволяющим ему распознавать жесты пользователя. Робот может взаимодействовать с пользователем в ходе игр, направленных на реабилитацию пациента. Так, одна из игр состоит в том, что робот пытается отобрать у пользователя некий предмет (например, маркер), тем самым побуждая последнего поднимать или опускать руки, т.е. совершать полезные с реабилитационной точки зрения движения. В ходе игры пользователь может давать устные указания роботу, облегчая взаимодействие.

Широкую известность приобрел антропоморфный робот KASPAR, разработанный в Хартфордширском университете (Великобритания) [14]. KASPARпозволяет детям с аутизмом развивать свои навыки социального взаимодействия. Представляет он собой антропоморфного робота размером с ребенка. Помимо развития навыков социального взаимодействия, робот также может использоваться при работе с детьми, страдающими речевыми дисфункциями.

Для полноты картины перечислим, не вдаваясь в подробности, еще несколько ассистивных роботизированных платформ: Ksera [15], Domeo [16], Cakna [17], Russell [18], Kirobomini [19].

Подавляющее большинство ассистивных роботов, рассмотренных или упомянутых в этом разделе, обладает речевым каналом обмена информацией и оснащаются комплексом устройств, необходимых для ухода за пользователем. Если последнее прямым образом проистекает из предназначения подобных роботизированных систем, то первое – наличие речевого канала – объясняется минимизацией усилий, необходимых для достижения максимальной эффективности взаимодействия пациента с роботом.

Роботы в образовательной сфере

Роботы достаточно широко используются и в образовательной сфере. Основными предметами, для обучения которым используются роботы, являются точные науки (информатика [20], математика [21]), прикладные науки (робототехника [22]), а также изучение иностранных языков. Роль робота в образовательном процессе варьируется от наглядного пособия до учителя: робот может быть наглядным пособием, учеником [23], помощником учителя или в некоторой степени – учителем [24].Учебные роботы весьма разнообразны по своим типам: от гуманоидных, таких как Asimo, до простых игрушечных конструкторов, например,LEGOMindstorms.

В качестве примера популярных промышленных моделей, использующихся в образовательной сфере, можно привести роботов Nao и DARwInOP.

Nao – перепрограммируемый робот от французской компании Aldebaran Robotics. Этот гуманоидный робот немногим более полуметра в высоту позиционируется как платформа для исследований в области робототехники или робот-учитель и находит реальное применение в сфере образования [25]. В 2014 году была выпущена последняя на сегодня модель Nao Evolution. с поддержкой синтеза и распознавания речи, локализацией источника звука, а также идентификацией лиц. Также Nao способен выполнять широкий диапазон движений, таких как ходьба, сидение, вставание, обход препятствий, удар ногой, захват объектов и танец.

Робот DARwIn реализован как модель с открытым кодом и модульной конфигурацией, что также позволяет использовать его в качестве образовательного робота в таких дисциплинах, как робототехника и программирование [26].

Роботы RUBI разработаны для работы в сфере раннего детского образования. Интересной особенностью проекта является то, что RUBI должен обладать хорошим запасом надежности, поскольку он должен тесно взаимодействовать с малышами.

Сам робот (последняя модель, RUBI-5) создан из готовых элементов, доступных в продаже, а информация по системе автоматического проектирования находится в открытом доступе [27].

Роботы Dash и Dot (в 2013-14 гг. известные как Bo и Yana) – обучающие роботы для детей и подростков от американской компании Wonder Workshop. Dash и Dot могут быть использованы для обучения ребенка основам программирования [28]. Разработчики считают, что их решение подходит уже для 5-летних детей. Роботы могут быть снабжены дополнительными аксессуарами (ковш, ксилофон, катапульта, буксировочный крюк и т.п.) для совершения различных действий и манипуляций, а также имеют модули для крепления элементов конструктора Lego.

Скачиваемые приложения позволяют создавать алгоритмы для роботов разной степени сложности: от простого перемещения по горизонтальной поверхности до воспроизведения музыки на ксилофоне, ориентации в лабиринте, расчистке территории. Работа с приложениями осуществляется посредством смартфона или планшета, интерфейс нагляден и понятен для ребенка.

Отдельного упоминания заслуживает ROBOBO – проект многомодального малобюджетного робота для проведения занятий в сфере точных и прикладных наук [29]. Основной акцент этой разработки лежит на выражении эмоций. Авторы разработки полагают, что, поскольку учебный процесс относится к социальной сфере человеческой деятельности, образовательный робот должен распознавать и выражать эмоции посредством мимики.

Ядром ROBOBO является смартфон, установленный на подвижную базу, оснащенную шасси и датчиками. Использование смартфона открывает доступ к таким технологиям, как синтез и распознавание речи, идентификация лиц, веб-сервисы для общения.

Существующие на сегодня гуманоидные роботы, заявленные как учителя, например, SAYA [30], несмотря на наличие многомодального интерфейса, не способны полноценно вести занятия; их функции сводятся преимущественно к презентации некоторой информации и проверке того, как она усвоена. Поэтому функции роботов в образовании сводятся, помимо телеприсутствия, к роли помощника и наглядного пособия, способного обмениваться информацией при помощи речевого канала.

Космическая сфера

Применение роботов в космической сфере связано с рядом специфических требований, что заставляет выделять этот класс роботов в отдельную категорию. Прежде всего, косморобот должен обладать высокой отказоустойчивостью и функционировать в условиях экстремальной среды. Это условие связано с большими сложностями, возникающими при необходимости ремонта или замены робота в условиях космоса, а также его высокой стоимостью.

Разумеется, к космороботам предъявляются повышенные требования по стабильности взаимодействия с пользователем. Зачастую общение проходит в отсутствие непосредственной близости и даже прямой видимости оператора. Важность стабильного человеко-машинного взаимодействия в условиях космической деятельности вынуждает вести разработки интерфейсов, поддерживающих не только «стандартные» речевую или сенсорную модальность.

Хорошим примером космического робота, взаимодействующего с пользователем при помощи «нестандартных» каналов обмена информацией, является проектMOONWALKпо разработке технологий для будущих космических миссий на Марс и Луну в Европейском космическомагентстве (ESA) был спроектирован и испытан многомодальный интерфейс для жестового управления роботом-ассистентом [31]. Основной причиной жестового управления является неудобство «традиционного» ручного управления в тяжелом скафандре. Очевидное, казалось бы, решение – голосовое управление – также оказывается неудобным и непрактичным из-за шума в скафандре, вызванного циркуляцией воздуха. Жестовое управление способно решить некоторые задачи по взаимодействию космонавта с роботизированным ассистентом. Для отслеживания жестов в проекте MOONWALK применяются инерционные блоки измерения, закрепленные на скафандре пользователя. Инерционные датчики снимают условие непосредственного визуального контакта между космонавтом и роботом, что может быть полезно в условиях затрудненной видимости, например, во время песчаной бури. Помимо жестового интерфейса, космонавт имеет доступ к сенсорным дисплеям, закрепленным на его скафандре в области груди. Также технологии, разработанные в рамках MOONWALK, подразумевают «узнавание» космонавта роботом посредством видеокамеры.

Прочие разработки в сфере космической робототехники, актуальные для настоящего обзора: проект «робозмеи»[32] и FEDOR [33].

Социальные роботы

К социальным роботам относятся роботы, разработанные для функционирования в социальной сфере: обслуживанию большого количества людей, предоставлению информации и навигации. Социальные роботы рассчитаны на взаимодействие с самыми разными слоями населения, поэтому их интерфейс должен быть интуитивно понятным и многомодальным. Среди прочих требований к социальным роботам можно перечислить «дружелюбную» внешность, мобильность и легкость настройки в зависимости от конкретных функций.

Робот-промоутер Promobot от одноименной российской компании позиционируется как сервисный робот для бизнеса, способный эффективно общаться и обслуживать многочисленных клиентов[34]. К основным особенностям Promobot следует отнести навигацию в местах повышенного скопления людей, понимание и синтез речи для ответов на интересующие пользователя вопросы, возможность вывода демонстрационных (рекламных) материалов на встроенный дисплей, распознавание лиц и распознавание по биометрическим данным.

В начале 2018-го года была выпущена обновленная версия v4, рассчитанная на массовое производство. Модели снабжены встроенными терминалами оплаты и функцией печати чеков.

HALA – робот-секретарь, разработанный в университете Карнеги Меллон, чьей отличительной особенностью является использование визуализированного на сенсорном мониторе 3D-аватара вместо роботизированной головы[35]. Благодаря этому система может взаимодействовать и со слабослышащими людьми. Движения губ аватара полностью совпадают с речевым сигналом, что дает возможность читать речь по губам. Помимо этого, робот оснащен микрофонами, динамиками и инфракрасным датчиком для определения присутствия пользователя. HALA может вести речевой диалог с пользователем на арабском и английском языках (в их литературных вариантах), при этом аватар позволяет использовать невербальные средства коммуникации, такие как покачивания головой или мимику.HALA используется, помимо всего прочего, для исследований в области человеко-машинного взаимодействия.

Cosero (Cognitive Service Robot) – сервисный робот, созданный в Боннском университете [36]. Основные задачи, поставленные при разработке этого сервисного робота, включают уверенную навигацию в помещении, манипулирование объектами и многомодальное взаимодействие с пользователем. Робот перемещается на восьмиколесном шасси, оснащен направленным микрофоном, лазерными сканерами, датчиками расстояния, двумя обычными и одной времяпролетной камерой. Также у робота имеется гуманоидная голова и два манипулятора, имитирующих человеческие руки.

Cosero способен понимать речевые команды, запоминать отдельных пользователей, следовать за человеком и создавать карту местности, он может, например, включить кухонную плиту или разбить яйца для приготовления омлета.

Прочие социальные роботы с многомодальным интерфейсом: Robotinho [37], Papero [38], Emiview-2 [39], Furo-S [40], REEM [41].

Анализ данных

Таблица 1 демонстрирует частоту применения того или иного канала обмена информацией с пользователем у ряда роботов, сгруппированных по сфере применения.

Абсолютное большинство рассмотренных интерфейсов являются многомодальными, поддерживающими распознавание (90% случаев) и синтез (80% случаев) речи, т.е. использующими речевой канал как для приема, так и для передачи информации. Во многом это связано с общей тенденцией перехода к парадигме интеллектуальных пространств, требующих естественных способов взаимодействия пользователя и искусственного интеллекта. Максимально привычной и естественной для современного человека формой коммуникации является речевая, очевидно, именно поэтому технологии синтеза и распознавания речи получили наибольшее распространение в многомодальных интерфейсах.

Таблица 1

Распределение роботизированных платформ в зависимости от набора модальностей

	Группа I	Группа II	Группа III	Группа IV	Группа V	Космороботы
синтез речи	+	+	+	+-	+	-+
распознавание речи	+	+	+	+	+	-+
распознавание лиц	+-			+-	+-	-+
сенсорный экран				+	-+	-+
манипулятор	+		+	-+	-+	-+
распознавание мимики	+			-+	-+	
имитация мимики	+-		+-			
распознавание жестов	+		+-	-+		+-
датчик прикосновений			+-			
аватар				+-	-+	

В Таблице 1 выделены основные группы роботизированных платформ, сгруппированные по количеству и сочетанию модальностей. К первой группе относятся роботы, обладающие широким спектром способов обмена информацией с пользователем: они способны, помимо голосо-речевого ввода, к распознаванию мимики, жестов, а также оснащены манипулятором. Отличительной особенностью является отсутствие сенсорного экрана, т.е. общение с таким роботом оказывается максимально приближенным к общению пользователя с другим человеком. Во вторую группу, наоборот, попали роботы, оснащенные «минималистичным» интерфейсом: единственным каналом обмена информацией с пользователем для них оказывается речевой. Признаком третьей группы является наличие речевого ввода и манипулятора. Дополнительно некоторые из роботизированных платформ этой группы обладают мимическим и жестовым каналом обмена информацией.

К четвертой группе относятся платформы, в которых совмещены речевой ввод и сенсорный. К пятой группе относятся интерфейсы, главной характеристикой которых является наличие речевого канала, совмещенного с еще одной модальностью: сенсорной, мимической, распознаванием лиц, 3D-аватаром и т.п. Космороботы, в силу специфики способов обмена информацией с пользователем, выделены в отдельную группу.

По-видимому, конкретная сфера применения сервисных роботов оказывается релевантной для выбора модальностей только в случае с космороботами. Распределение остальных платформ по вышеперечисленным группам относительно равномерно. Можно говорить о том, что сервисные роботы с многомодальным интерфейсом решают похожий спектр задач в разных областях, поэтому невозможно выделить характерный набор модальностей, скажем, для образовательной сферы или ассистивной.

Скорее, речь идет о том, что каждая конкретная платформа разрабатывается под конкретные задачи, это и определяет выбор той или иной модальности. Это не значит, что все модальности равноправны; скорее, наоборот, общей тенденцией является поддержка распознавания и синтеза речи (90% рассмотренных интерфейсов), зачастую – в комбинации с графическим (сенсорным) дисплеем (40%). Прочие модальности используются только по мере необходимости; так, для робота-секретаря естественно обладать функцией распознавания и запоминания лиц, для роботов, используемых для реабилитации пациентов – распознавание эмоций и т.п.Можно сделать вывод о том, что развитие многомодальных интерфейсов идет не по пути наибольшей естественности, а наибольшей простоты и минимизации усилий со стороны пользователя. Голосовой ввод во всех отношениях проще сенсорного или графического, однако естественность взаимодействия достигается при использовании тех каналов обмена информацией, к которым прибегает пользователь в зависимости от коммуникативной ситуации.

Заключение

Итак, основными областями внедрения роботизированных платформ с многомодальным интерфейсом оказываются ассистивная, образовательная и социальная, причем в последнем случае главными функциями робота являются информационная и развлекательная (промоботы).

Между тем, применение многомодальных технологий в сфере сервисной профессиональной робототехники представляется вполне оправданным. Хорошим примером являются промоботы: их использование указывает на то, что спрос на «роботизированных помощников» и предоставляемые ими услуги по навигации и информированию в местах повышенного скопления людей достаточно велик. Следующим шагом по развитию этого сегмента рынка роботов может быть совмещение услуг информационного характера и выполнения полезной работы. Именно таким набором функций обладает роботизированная тележка для супермаркетов – проект авторов данного обзора.

Идея роботизированной платформы, которая перевозит личные вещи пользователя, разумеется, не нова. Относительно недавно были созданы робот-носильщик Gita от итальянского концерна Piaggio [42] и роботизированная тележка Dash от FiveElementsRobotics [43], однако ни один из существующих роботов-носильщиков не обладает по-настоящему многомодальным интерфейсом. В СПИИРАН с 2018 г. выполняется проект, ориентированный на разработку многомодального интерфейса для мобильной тележки для продуктового супермаркета, обладающей следующими особенностями:

1. понимание голосовых и жестовых команд;
2. понимание команд на РЖЯ;
3. наличие сенсорного экрана;
4. сопровождение пользователя до витрины с интересующим ему товаром, до определенного места в магазине (касса, выход, туалет)
5. режим «followme»;
6. синтез речи;
7. наличие 3D-аватара для синтеза ответов на русском жестовом языке;
8. полезная нагрузка до 12 кг.

Хорошо известно, что навигация по супермаркету в поисках нужного товара не всегда оказывается легким делом для покупателя. Зачастую только на поиск нужного отдела уходит большая часть времени, проведенного в магазине. Внедрение роботизированной тележки вполне решает эту проблему, сберегая пользователю время и силы. Применение многомодальных технологий, в частности, речевого и жестового каналов обмена информацией, практически полностью освобождает покупателю руки, делая взаимодействие с роботом легким и ненавязчивым, а понимание жестового языка позволяет пользоваться тележкой даже слабослышащим людям. Напомним, что общее число носителей русского жестового языка на 2010 год составляло 122 тысячу человек, из которых 121 тысяча проживали на территории России [44]. В 2012-м году русский жестовый язык был признан языком общения на территории Российской Федерации «при наличии нарушений слуха и (или) речи, в том числе в сферах устного использования государственного языка» [45].

Использование роботизированной тележки с многомодальным интерфейсом имеет следующие преимущества:

1. Быстрота поиска товара приводит к увеличению объема продаж;
2. быстрота поиска товара и легкость навигации привлекут больше покупателей;
3. роботизированная тележка является платформой, позволяющей покупателю чувствовать себя мобильнее: на тележку могут быть установлены прайс-чекеры, устройства оплаты банковской картой и т.п.
4. роботизированная тележка повышает качество жизни слабослышащих людей, функция понимания русского жестового языка позволяет им пользоваться тележкой наряду с остальными клиентами магазина.

Данное исследование проводится при поддержке Министерства науки и высшего образования РФ (Проект «Многомодальный интерфейс на основе жестов и речи для управления ассистивным мобильным информационным роботом» по Соглашению № 14.616.21.0095, идентификатор соглашения RFMEFI61618X0095).

ЛИТЕРАТУРА

1. World Robotics – Industrial robots 2017: Statistics, Market Analysis, Forecasts and Case Studies. Frankfurt-am-Main: VDMA Verlag, 2017.
2. The Global Service Robotics Market. Göteborg: Berg Insight AB, 2016.
3. World Robotics – Service robots 2017: Statistics, Market Analysis, Forecasts and Case Studies. Frankfurt-am-Main: VDMA Verlag, 2017.
4. Карпов А.А., Юсупов Р.М. Многомодальные интерфейсы человеко-машинного взаимодействия // Вестник Российской академии наук. – 2018. – Т. 88(2). – С. 146–155.
5. Ронжин А.Л., Карпов А.А., Ли И.В. Речевой и многомодальный интерфейсы. М.: Наука, 2006.
6. Ронжин А.Л., Юсупов Р.М. Многомодальные интерфейсы автономных мобильных робототехнических комплексов // Известия Южного федерального университета. Технические науки. – 2015. – № 1(162). – С. 195–206.
7. Кагиров И.А., Карпов А.А. Многомодальные интерфейсы для сервисных роботов (аналитический обзор) // Пилотируемые полеты в космос. — 2018. — №4. (в печати).
8. Toyota Partner Robot // [Электронный ресурс] – Электронные данные – Режим доступа: URL: http://www.toyota-global.com/innovation/partner_robot/family_2.html#h210 свободный.
9. Companionable research project delivers robotic assistance for the elderly // [Электронныйресурс] – Электронные-данные – Режимдоступа: URL: https://ec.europa.eu/digital-single-market/en/news/companionable-research-project-delivers-robotic-assistance-elderly свободный.
10. KOMPAÏ: TheConnectedHealthcareAssistant // [Электронный ресурс] – Электронные данные – Режим доступа: URL: https://kompai.com/docs/kompaiflyer_en.pdf свободный.
11. AAL Programme: ALIAS // [Электронный ресурс] – Электронные данные – Режим доступа: URL: http://www.aal-europe.eu/projects/alias/ свободный.
12. Calderita L., Bustos P., Suárez Mejías C., Fernández F., Bandera A. THERAPIST: Towards an autonomous socially interactive robot for motor and neurorehabilitation therapies for children // Proceedings of the 7th International Conference Pervasive Computing Technologies for Healthcare and Workshops. Venice, 2013. pp. 374–377.
13. Dautenhahn K., Nehaniv Ch., Walters M., Robins B., Kose H., Assif Mirza N., Blow M. (2009). KASPAR – A Minimally Expressive Humanoid Robot for Human–Robot Interaction Research // Applied Bionics and Biomechanics. 6. 2009. pp. 369–397.
14. Mayer P., Beck Ch., Panek P. (2012). Examples of multimodal user interfaces for socially assistive robots in Ambient Assisted Living environments // 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Kosice 2012. pp. 401–406.
15. TheDOMEOHomepage // [Электронный ресурс] – Электронные данные – Режим доступа: URL: http://www.aat.tuwien.ac.at/domeo/index_en.html свободный.
16. Aziz A.A., Saad A., Ahmad F. CAKNA: A Personalized Robot-based Platform for Anxiety States Therapy // Proceedings of the 13th International Conference of Intelligent Environments. Seoul, 2017. pp.141–150.
17. Humanoid Robot Russell Enganes Children with Autism // [Электронныйресурс] – Электронныеданные – Режимдоступа: URL: https://dcmp.org/media/9089-humanoid-robot-russell-engages-children-with-autismсвободный.
18. The small robot with big possibilities: Introducing Toyota's Kirobo Mini // [Электронный ресурс] – Электронные данные – Режим доступа: URL: https://www.toyota-europe.com/world-of-toyota/articles-news-events/introducing-kirobo-minicвободный.
19. Balch T., Summet J., Blank D., Kumar D., et al. Designing personal robots for education: hardware, software, and curriculum // IEEE, Pervasive Computing. 2008. vol. 7(2). pp. 5–9.
20. Highfield K., Mulligan J., Hedberg, J. Early mathematics learning through exploration with programmable toys // Proc. Joint Conference of the International Group for the Psychology of Mathematics Education (32nd: 2008) and the North American Chapter of the Psychology of Mathematics Education (30th: 2008), Morelia, Mexico 2008. pp. 169–176.
21. Chiou A. Teaching technology using educational robotics // UniServe Science Scholarly Inquiry Symposium Proceedings 2004. pp. 9–14.
22. Okita S.Y., Ng-Thow-Hing V., Sarvadevabhatla R. Learning together: asimo developing an interactive learning partnership with children // Proc. RO-MAN. 2009 – The 18th IEEE International Symposium on Robot and Human Interactive Communication, Toyama, 2009. pp. 1125–1130.
23. Goodrich M.A., Schultz A. Human robot interaction: a survey // Foundations and Trends in Human-Computer Interaction. – 2007. – vol. 1(3). – pp. 203–275.
24. Majgaard G. Multimodal Robots as Educational Tools in Primary and Lower Secondary Education // The 9th International Conference on Interfaces and Human Computer Interaction. IADIS Multi Conference on Computer Science and Information Systems, 22-24 July 2015. pp. 27–34.
25. Роботы DARwIn // [Электронный ресурс] – Электронные данные – Режим доступа: URL: http://robotgeeks.ru/collection/darwin свободный.
26. Johnson D., Malmir M., Forster D., Alac M., Movellan J. Design and early evaluation of the RUBI-5 sociable robots // Proceedings of IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL). San Diego, CA, 2012. pp. 1–2.
27. Dashisachild'sfirstrealrobotfriend // [Электронный ресурс] – Электронные данные – Режим доступа: URL: https://www.makewonder.com/dash свободный.
28. Naya M., et al., A versatile robotic platform for educational interaction // 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). Bucharest, 2017. pp. 138–144.

29. RobottakesoverTokyoclassroom // [Электронный ресурс] – Электронные данные – Режим доступа: URL: https://www.reuters.com/article/us-japan-robot-teacher-life/robot-takes-over-tokyo-classroom-idUSTRE54A69K20090511 свободный.
30. Weiss P. et al. The Moonwalk Project Astronaut-Robot Cooperation in European Space Analogue Simulations // [Электронный ресурс] – Электронные данные – Режим доступа: URL: http://www.projectmoonwalk.net/moonwalk/?p=781 свободный.
31. Jinguo Liu, Yifan Luo, Zhaojie Ju. An Interactive Astronaut-Robot System with Gesture Control // Computational Intelligence and Neuroscience. – 2016. – vol. 2016. – p. 11.
32. Робот Федор станет героем России // [Электронный ресурс] – Электронные данные – Режим доступа: URL: https://lenta.ru/news/2017/11/20/became_an_hero/ свободный.
33. Встречайте новый PROMOBOT V.4 // [Электронный ресурс] – Электронные данные – Режим доступа: URL: https://promo-bot.ru свободный.
34. Makatchev M., et al. Dialogue patterns of an Arabic robot receptionist. // 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Osaka 2010. pp. 167–168.
35. Stückler J., Schwarz M., Behnke S. Mobile Manipulation, Tool Use, and Intuitive Interaction for Cognitive Service Robot Cosero // Frontiers in Robotics. – 2016 – vol. 3.
36. Faber F., et al. The humanoid museum tour guide Robotinho // The 18th IEEE International Symposium on Robot and Human Interactive Communication. Toyama, 2009. pp. 891–896.
37. Японский робот-компаньон стал настольным // [Электронный ресурс] – Электронные данные – Режим доступа: URL: https://lenta.ru/news/2013/11/15/papero свободный.
38. Nakamura R., et al. Development of Human-symbiotic Robot EMIEW2: Mechanism and System Constitution // Proceedings of ROBOMEC 2008. pp. 1–3.
39. FutureRobot: FURo-Desk // [Электронный ресурс] – Электронные данные – Режим доступа: URL: http://www.futurerobot.co.kr/en/page/product01.php свободный.
40. Tellez R., et al. Reem-B: An autonomous lightweight human-size humanoid robot // Humanoids 2008 - 8th IEEE-RAS International Conference on Humanoid Robots. Daejeon. 2008. pp. 462–468.
41. IntroducingGita // [Электронный ресурс] – Электронные данные – Режим доступа: URL: https://www.piaggiofastforward.com/gita свободный.
42. DashRetailRoboticsShoppingCart // [Электронный ресурс] – Электронные данные – Режим доступа: URL: http://5elementsrobotics.com/dash-retail-robotic-shopping-cart/ свободный.
43. RussianSignLanguage // [Электронный ресурс] – Электронные данные – Режим доступа: URL: https://www.ethnologue.com/language/rsl свободный.
44. Федеральный закон от 30 декабря 2012 г. N 296-ФЗ «О внесении изменений в статьи 14 и 19 Федерального закона «О социальной защите инвалидов в Российской Федерации» // [Электронный ресурс] – Электронные данные – Режим доступа: URL: http://base.garant.ru/70291470 свободный.

I.A. Kagirov (SPIIRAS, St. Petersburg).

Analytical overview of robotics systems with multimodal user interfaces.

The present article proposes an analytical review of robotic platforms with multimodal interfaces. Some trends in different ways of HCI are analysed depending on the area of robot application.The main trend is the minimization of efforts on the part of the user. A project of a robotic trolley supporting exchange of information with the user via the voice channel is proposed, as well as communication with the help of the Russian sign language.

ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ ПРОСТРАНСТВ ДЛЯ ИНФОРМАЦИОННОЙ ПОДДЕРЖКИ ОПЕРАТОРА ПРИ ДИСТАНЦИОННОМ МОНИТОРИНГЕ ГРУППЫ МОБИЛЬНЫХ РОБОТОВ НА ЛУННОЙ ПОВЕРХНОСТИ

Б. И. Крючков, **В. М. Усов**
(*ФГБУ «Научно-исследовательский испытательный центр подготовки космонавтов имени Ю.А. Гагарина», 141160, Звездный городок Моск. обл.,*
тел: +7 917 550-4807, b.kyuchkov@gctc.ru;
тел: +7 910 459-4902, v.usov@gctc.ru)

Д. В. Иванько
(*ФГБУН «Санкт-Петербургский институт информатики и автоматизации Российской академии наук», 199178, С.-Петербург, 14 линия, 39,*
тел. +7 921 362-6287, denis.ivanko11@gmail.com)

Аннотация

Рассматривается проблема построения онтологий совместного решения задач человеком и автономными мобильными роботами (АМР) в экстремальной среде. Показано, что приложение концепции интеллектуального пространства к управлению задачами группы автономных мобильных роботов (АМР) при внекорабельной деятельности (ВКД) на лунной поверхности требует распределения функций в человеко-машинной системе «космонавт – группа АМР», исходя из антропоцентрического принципа ее организации. Для мониторинга человеком-оператором (ЧО) текущей ситуации, исходя из основного критерия – безаварийного функционирования группы АМР, предлагаются варианты построения интеллектуального пространства, позволяющие обеспечить ситуационную осведомленность ЧО и реализацию его своевременного вмешательства в управление группой АМР посредством активации, приостановки и/или прекращения отдельных задач (упорядоченных по времени согласно структуре полетных операций и контролируемых ЧО с точки зрения конфликтов по заявкам на ресурсы). Для принятия решений космонавтом о предотвращении возможных коллизий участников ВКД и/или конфликтов интересов на уровне учета и планирования ресурсов показана необходимость построения интерактивной виртуальной среды, обеспечивающей моделирование и визуализацию текущей ситуации на основе совокупности данных, получаемых от активных интеллектуальных агентов, в данном случае от группы АМР, предварительно составленных электронных карт и планов выполнения операций с помощью группы АМР. На этой основе возможен синтез виртуального окружения и отображение взаимодействующих в ней виртуальных агентов с индикацией областей потенциальных коллизий для оперативного привлечения внимания ЧО и его включения в предотвращение коллизий. Данное решение учитывает роль человеческого фактора в обеспечении безопасности ВКД при возникновении потенциально опасных ситуации. Применение АМР для интерактивного сбора информации об экстремальной среде способствует более полному формированию контекста принятия решений, и в этом смысле расширяет традиционную трактовку понятия интеллектуального пространства.

Ключевые слова: человек-оператор (ЧО); группа автономных мобильных роботов (АМР); мониторинг АМР на лунной поверхности; интеллектуальное пространство.

Введение

В настоящее время активно обсуждаются варианты определения понятия «интеллектуального пространства» (*англ. синонимы*: Smart Space – SS, Intelligent Environments – IE, Augmenting Smart Space – ASS), а также перспективы его использования в различных прикладных областях. В этой связи предлагаются различные варианты построения моделирующих систем. В зарубежной литературе смысловое наполнение этого понятия варьируется в зависимости от контекста, в частности, исходя из ролевой позиции человека. В общем случае интеллектуальное пространство представляет собой инфраструктуру для взаимодействия сервисов, описывающих физические устройства, которые совместно решают запланированные и вновь возникающие задачи с учетом присутствия человека и отслеживания контекста ситуации. В тех случаях, когда человек выступает как потребитель определенного перечня услуг и сервисов, акцент на расширение автономности киберфизических систем, осуществляющих сбор информации о самом человеке, его потребностях или же о параметрах условий, оказывающих влияние на состояние человека [1, 2].

В соответствии с этой трактовкой цель интеллектуального пространства состоит в оказании человеку «осмысленной помощи», даже без его участия. В отечественной литературе также описаны результаты проектирования прикладных систем (такие как «умный дом», «умный автомобиль», «умный город») [3, 4], основывающихся на указанной трактовке интеллектуальных пространств.

В некоторых, более сложных вариантах анализа потенциальной применимости понятия Intelligent Environments (IE), предлагается рассматривать человека (оператора) как активное звено взаимодействия с интеллектуальными агентами, совместно выполняющими сложный набор прикладных задач [5].

Для этого случая подходит термин «окружающего интеллектуального пространства» (*англ.*: Augmenting Smart Space – ASS). Дополнительно для расширения возможностей ситуационной осведомленности человека-оператора (ЧО), могут быть использованы интеллектуальные роботы, имеющие развитые сенсорные возможности и средства коммуникации. Следует заметить, что при этом умственная нагрузка на ЧО может существенно возрастать из-за необходимости контроля активности интеллектуальных агентов в сложно структурированной среде.

В работе [5] предлагается использовать серийные роботы, чтобы они могли:

1) выступать в качестве физического средства сбора данных для интеллектуального пространства;
2) обеспечивать взаимодействие между пользователями на постоянной основе;
3) предоставлять человеку в определенном объеме «ассистивные услуги»;
4) позволять осуществлять захват дополнительной «контекстной» информации, которую трудно или невозможно зафиксировать с помощью стационарных датчиков.

Таким образом, посредством применения (интеллектуальных) роботов предлагается улучшить взаимодействие повсеместной вычислительная инфраструктуры с физической средой.

При рассмотрении любых вариантов построения интеллектуального пространства необходимо учитывать человеческий фактор, когда остро стоят вопросы обеспечения безопасности и безаварийности [6-9].

Еще одним чрезвычайно важным направлением отечественных разработок является исследование многомодальных интерфейсов при определении функциональности интеллектуальных пространств [10, 11].

Исходя из существующих общеметодических подходов, далее будет рассмотрена предметная область пилотируемой космонавтики, а, более конкретно, – перспективы ее развития в отношении выполнения ВКД на лунной поверхности с использованием космических роботов.

Особенности взаимодействия человека и робота на уровне совместно решаемых задач ВКД на лунной поверхности

Для дальнейшего изложения необходимо дать характеристику тех проектов экстремальной робототехники, которые связаны с лунной программой. Как известно, линия на продолжение исследования космического пространства автоматами активно поддерживается рядом исследовательских групп, но и в пилотируемой космонавтике эти технические решения получают реализацию на уровне проектов. На основе накопленного опыта и анализа предлагаемых проектных решений многие исследователи выдвигают свои версии, как будут формироваться программы полетов и какие особенности деятельности экипажа космонавтов в экстремальных условиях необходимо учитывать для успеха экспедиций [12-15].

В числе ожидаемых видов деятельности космонавтов во время напланетной ВКД при выполнении лунных миссий называются следующие [16, 17]:

- решение целевых задач (выполнение комплексной программы научных исследований и экспериментов; отработка новых космических технологий, добыча и переработка лунных ресурсов);
- участие в создании (построении, сборке) космической инфраструктуры (на Луне);
- работа с доставляемым оборудованием (разгрузочно-погрузочные работы, инвентаризация доставляемых и удаляемых грузов, транспортировка по поверхности Луны);
- управление робототехническими комплексами (орбитальными, напланетными, в т.ч. антропоморфного типа);

− обеспечение безопасности напланетной деятельности и оказание медицинской помощи пострадавшему космонавту в случае медицинской нештатной ситуации или инцидента, приведшего к потере работоспособности, в том числе, отказе средств защиты и систем обеспечения жизнедеятельности.

Например, проведение тематической геологической съемки и предварительной геологической разведки, согласно [17], является важным и необходимым этапом освоения Луны.

На этом этапе предполагается не только собрать обширную научную информацию о самой Луне, но и конкретные исходные данные и сведения, необходимые для размещения долговременной или периодически посещаемой космонавтами лунной базы, что является прологом к развертыванию «первичной промышленной инфраструктуры».

В целом, можно полагать, что при освоении Луны будут опробованы различные модификации автономных мобильных роботов с разным функционалом, а экипаж будет располагаться на стационарной лунной базе и обеспечивать безопасность работ на основе дистанционного мониторинга перемещений роботов.

Таким образом, появляется ранее не исследованная задача в составе ВКД космонавтов. При этом прогнозируемый вид операторской деятельности можно отнести к разновидности диспетчерских профессий, дополненных оперативным планированием и анализом предпосылок к инцидентам, связанных с потенциальными коллизиями и проблемами совместного функционирования группы роботов, имеющих различное предназначение.

Выполнение жестких требований по безопасности должно обеспечиваться прежде всего оснащенностью самих АМР для адаптации к внешней среде и согласованности в отношении безаварийного функционирования в соответствие с распределенными между ними задачами. В частности, необходимо выставлять требования к возможностям АМР выполнения навигации, трекинга подвижных объектов, выполнения дальнометрии, ведения связи и поддержания коммуникаций с тем, чтобы сформировать виртуальную среду принятия ЧО решений в интеллектуальном пространстве

При указанных выше предположениях об оснащении активных агентов из литературы известны разные решения по построению интеллектуального пространства. Например, в работах [9, 18,19] представлены подходы к взаимодействию людей и роботов для совместного решения задач, построенных на разных принципах разделения функций и компетенций, рассмотрены базовые и дополнительные функции роботов, которые могут быть использованы при разных подходах и показано, какие онтологические модель служат для организации знаний.

Построение человеко-машинного интерфейса для взаимодействия космонавтов и роботов при решении задач ВКД и среда моделирования виртуального окружения и визуализации для принятия решений ЧО в сложной ситуации

В литературе по проблемам построения диалога человека и робота, как формы взаимодействия «человек – роботы» (*англ.*: Human Robot Interaction - HRI) [20], этот круг методологических проблем рассматривается с позиций нахождения способов порождения единого семантического базиса (например, в форме онтологии), доступного для выражения привычных для человека категорий пространства и времени по отношению к объектам внешней среды в категориях «модели внешнего мира» робота.

В [21] рассмотрены возможные способы построения АМР, способных осуществлять автономную навигацию в недетерминированной рабочей среде.

При этом деятельность ЧО в значительной степени сводится к постановке текущих задач для роботов и установлению различных регламентов и ограничений. Значительную роль играет навигационная система самих АМР для ориентировки в пространстве, так как роботы должны самостоятельно оценивать окружающую среду и планировать свой путь, в том числе при наличии других движущихся объектов в контролируемой зоне. Информационное обеспечение деятельности ЧО должно учитывать особенности исходных данных при подготовке и принятии решений. Одним из путей решения данной проблемы является использование методов нечеткой логики на этапах планирования действий и оперативного принятия решений при возрастающем риске возникновения конфликтов в гетерогенной группе АМР.

Рассмотрим несколько тактических задач, которые на уровне ментальных представлений ЧО выступают как когнитивные составляющие принятия решения, а в рамках разрабатываемой онтологии задают связи между роботом, объектом воздействия, видом активности робота, допустимой областью активности в контролируемой территориальной области.

Наиболее простой вариант состоит в планировании размещения во времени и/или пространстве АМР так, чтобы маршруты их передвижения лежали в непересекающихся областях территориальной зоны.

Более сложный вариант представлен ситуацией, когда АМР периодически передвигаются по пересекающимися маршрутам. Можно рассмотреть два варианта реагирования на такие ситуации. Наиболее естественная тактика состоит в назначении одного из АМР лидером, который организует передвижение по своей зоне контроля других АМР. Такие постановки известны из работ, посвящённых проблеме самоорганизации роботов в группе. Альтернативой может служить тактика временных запретов для передвижения в смежных пересекающихся областях, когда ЧО вводит искусственный виртуальный объект – препятствие, которое возникает при сближении АМР на величину, опасную с точки зрения коллизий с учетом параметров их движения (направления и скорости). Также в качестве критериев для безопасного прохождения препятствия могут быть кинематические и динамические параметры мобильных роботов. Такие идеи можно найти в работах [22-25].

Чтобы обеспечить ситуационную осведомленность ЧО в режиме дистанционного мониторинга группы АМР, необходимо предусматривать наличие программно-информационных средств, позволяющих поддерживать в актуальном состоянии сведения, необходимые для планирования задач для группы АМР и контроля хода их выполнения. Этим обеспечивается главная цель, которая стоит перед ЧО, – добиться согласованного, бесконфликтного функционирования неоднородной группы АМР.

В качестве первой компоненты выступает распределенный программно-аппаратный комплекс сбора исходных данных от АМР, реализованный: 1) в составе бортовых устройств, которыми оснащаются АМР для ориентировки в среде, навигации, идентификации объектов; 2) в системе телеметрии и связи для обмена данными между АМР и для их аккумуляции в системе информационного обеспечения ЧО на рабочем месте стационарной лунной базы.

В качестве второй компоненты можно выделить средства, которые должны быть в распоряжении ЧО для интеграции и визуализации на системе отображения информации в доступной для восприятия форме всей совокупности данных, оперативно получаемых по каналам связи с АМР и доставляемых для интеграции на средства отображения информации на рабочем месте ЧО.

Предназначение третьей компоненты – поддержание механизма управления группой посредством постановки задач для АМР в рамках предлагаемой онтологии, и формирования на этой основе предписаний по составу цепочек выполняемых АМР задач.

В конечном итоге необходимо необходимо найти такую реализацию HRI, которая позволяет построить и визуальный образ расположения АРМ на контролируемой территории, и индикаторы удовлетворения запросов АМР на требуемые для их функционирования и перемещения ресурсы. Ключевой вопрос заключается в том, каким образом построить систему описания ситуации. Возможные принципы и методы построения среды, которая отвечает целям моделирования такой сложной системы, обсуждаются в ряде работ, краткий обзор которых приведен в работе [9].

Например, в [26] рассматриваются вопросы построения взаимодействия людей и роботов и улучшения ситуационной осведомленности ЧО на базе агентно-ориентированного подхода на примере управления передвижением смешанной команды АМР и группы космонавтов в защитном снаряжении по поверхности Луны на контролируемой территории.

Фактически необходимо различать несколько разных фаз во взаимодействии человека и роботов, отдельно рассматривая: 1) фазу сбора и доставки исходных данных для человека с учетом его компетенций и специализации; 2) фазу обработки и представления информации для человека с использованием электронных приборов и много модальных интерфейсов; 3) фазу подготовки и принятия решений о необходимости вмешательства человека в активность роботов; 4) фазу формирования задач в соответствие с разработанной онтологией.

На рис.1 представлена схема построения человеко-машинного интерфейса применительно к взаимодействию «человек – робот» (*англ.*: Human Robot interaction – HRI) при мониторинге ЧО группы АРМ.



Рис. 1. Построение HRI для взаимодействия «человек – робот» при мониторинге группы АРМ

Важной для принятия решений является компонента информационной поддержки на основе представления визуальной информации для космонавта, выполняющего функции диспетчера-наблюдателя, в распоряжении которого имеется сложный семантический конструкт, а именно многослойная «цифровая карта», допускающая различные формы и варианты ее отображения ЧО.

Важно подчеркнуть, что графический интерфейс типа «карта» должен не только представлять в наглядной форме текущее пространственное распределение активных агентов (людей и роботов) в виртуальном окружении, но и давать возможность отображать статус и информацию о событиях, полученных от каждого агента.

Это позволяет ЧО перейти на уровень управления задачами, используя их спецификации и текущие данные о статусе исполнения. В этих предположениях на рис. 2 показан фрагмент онтологии, ориентированной на построение интеллектуального пространства и отражающей специфику взаимодействия человека и роботов в ходе мониторинга активности группы АМР.

Краткое обоснование такого методического подхода базируется на предположении, что в отличие от способности решения роботами отдельных и, как правило, частных задач, человеку присуща способность понимать текущую ситуацию в целом.

Если операция, которая представлена цепочкой задач, каждая из которых характеризуется исполнителем, местом и временем, а также имеет четко описанные критерии успешного или неуспешного завершения, не пересекается с другими операциями, то диаграмма переходов состояния задачи можно представить последовательностью (Табл.1).

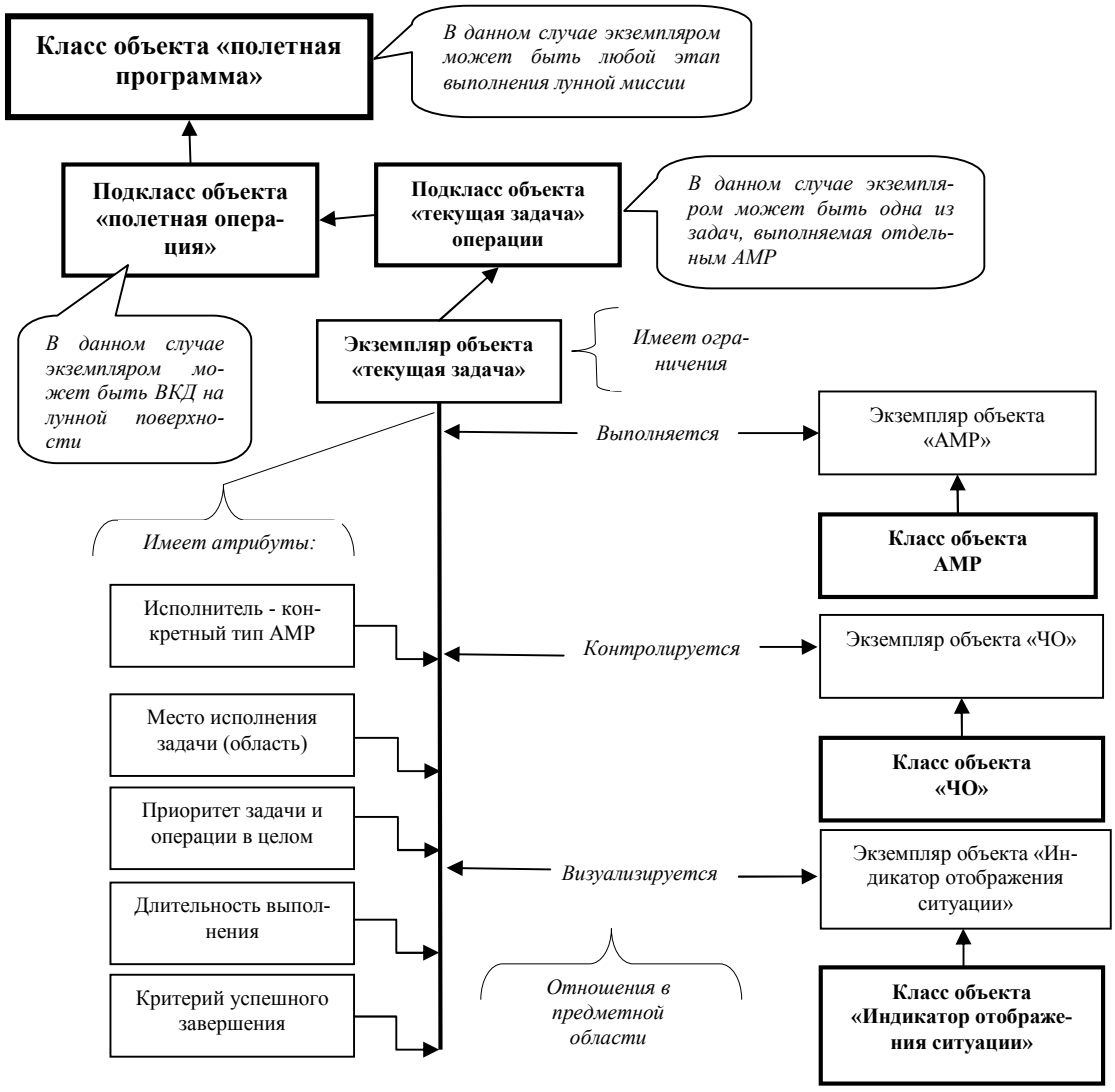


Рис. 2. Фрагмент онтологии «текущая задача в составе операций», выполняемых АМР

Таблица 1

Диаграмма переходов состояния исполняемой задачи для выполняемых параллельно операций для описания действий ЧО при конфликте интересов задач по выделяемым ресурсам

№№ п/п	Состояние задачи в последовательно исполняемой цепочке согласно составу операции и сценарию работ	сценарию работ
1	Ожидание завершения предшествующей в цепочке задачи	Автомат
2	Инициирование новой задачи с учетом исходных данных и придание ей статуса текущей с резервированием для нее запрашиваемых ресурсов	Автомат
3	Выполнение за заданный отрезок времени с откреплением приданных ресурсов	Автомат
4	Выработка признака успешного/неуспешного завершения	Автомат
5	Приостановка или завершение	Ожидание реакции ЧО

При успешном завершении осуществляются автоматические переходы по цепочке согласно сценарию, и вмешательства ЧО не требуется. При неуспешном завершении возможны тактические решения о возможности повторной попытке или же приостановке операции в целом в зависимости от того, используются ли результаты при выполнении последующих задач на усмотрение ЧО.

Значительно сложнее диаграммы переходов, когда выполняется несколько операций и возможны конфликты интересов по запрашиваемым ресурсам (Табл.2). В отношении «места исполнения» этот конфликт может быть источником коллизий между АМР, а потому требуется принятие решения о вмешательстве ЧО на уровне приостановки какой-то из «конфликтующих задач» или ее завершении с повторным исполнением, если это допустимо по технологическим соображениям.

Условный пример такой ситуации потенциального конфликта интересов, который может привести к коллизии АРМ, например в отношении места локализации выполнения задач, представлен на рис.3. в предположениях о более сложных соотношениях исполнения задач.

Необходимо отметить, что переход к управлению «на уровне задач», как способ обеспечения безаварийного взаимодействия в группе АРМ с позиций эргономики означает переход к способам индикации текущих ситуаций в терминах технологических операций и контроля в оперативном режиме функционирования сложной организационно-технической системе. Для такого варианта предусмотрены специальные виды человеко-машинного интерфейса, что требует отдельного анализа, ввиду специфики рассматриваемой предметной области и трудностей переноса принятых методов и средств контроля исполнении планов для орбитальных комплексов.

Таблица 2

Диаграмма переходов состояния исполняемой задачи для выполняемых параллельно операций для описания действий ЧО при конфликте интересов задач по выделяемым ресурсам

Состояние задачи	Управляющий признак	Необходимость вмешательства ЧО	Действия ЧО
Ожидание завершение предшествующей в цепочке задачи	Успех	Нет	Мониторинг
	Неуспех	Да	Повторное инициирование или завершение предшествующей
Инициирование новой задачи с учетом конфликта интересов с текущей задачей другой операции.	- нет	Нет	Мониторинг
	- есть	Да	Выбор задачи для приостановки или завершения конкурирующих задач и операций
Выполнение новой задачи, как текущей	Успех	Нет	Мониторинг
Завершение текущей задачи в штатном режиме	Успех	Да	Мониторинг
Нештатное завершение текущей задачи	Неуспех	Нет	Повторное инициирование или завершение операции

Представленное распределение функций в штатном и нештатном исполнении задач АМР в зависимости от выявляемых конфликтов интересов, которые ввиду возможных коллизий могут повлиять на безаварийное управление группой АМР, требует от разработчиков интерфейсов обратить особое внимание на способы активации когнитивных функций ЧО, который должен принимать решения в сложной тактической обстановке, используя виртуальные представления об активных агентах в моделируемой интерактивной индуцированной среде.

В настоящее время не представляется возможным достичь в полном объеме автоматизации, которая реализовала бы определенные когнитивные функции, присущие адаптивному поведению человека в проблемных ситуациях. Таким образом, проблема переводится в плоскость разделения функций на уровне планирования и исполнения задач в составе полетных операций, а человеку отводится роль «наблюдателя – диспетчера» в системе мониторинга активности неоднородной группы агентов. Это означает необходимость построения единой информационной базы той области, в которой функционирует группа (электронной карты), использования средств локализации всех участников в конкретные моменты времени, разработки средств введения ограничений по мобильности, имея в виду разнесения маршрутов перемещений по пространственно-временным параметрам и др.

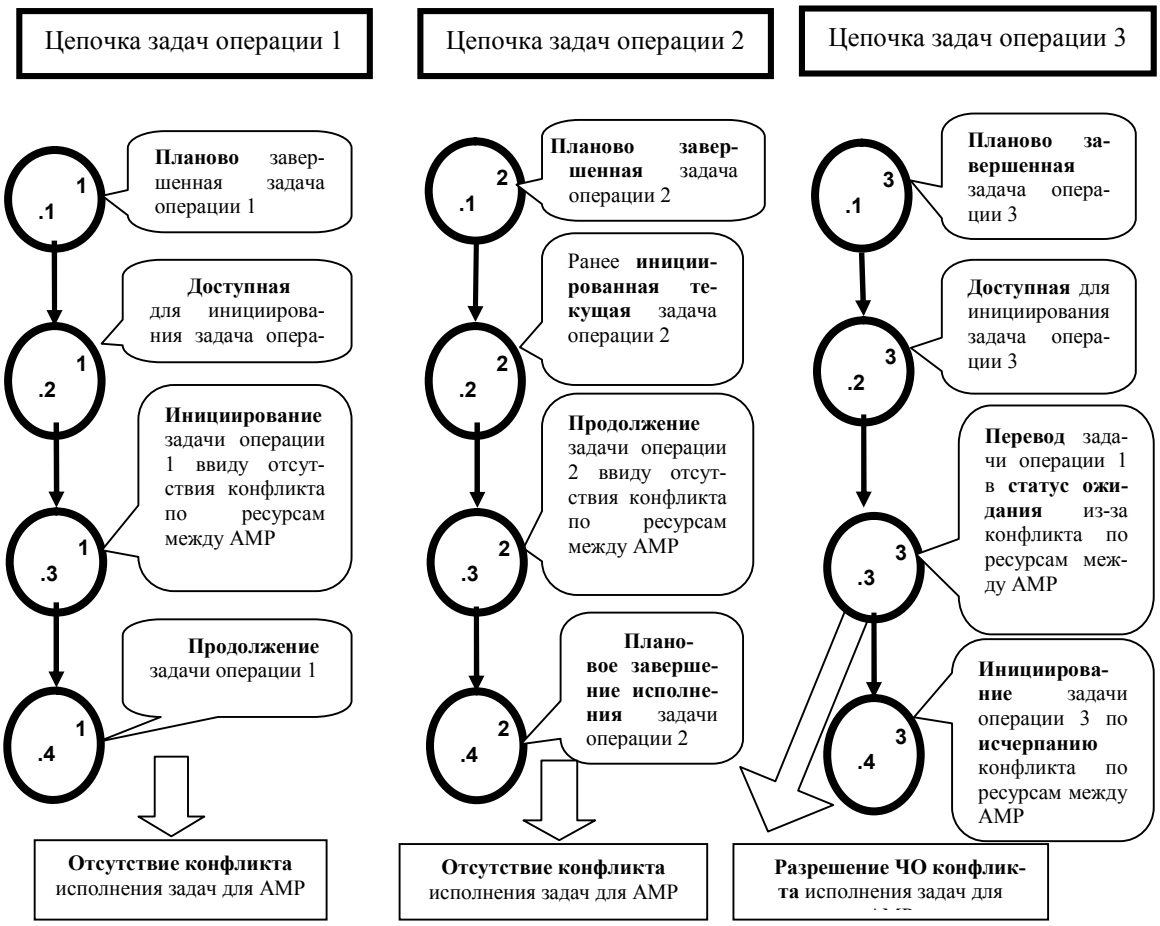


Рис. 3. Диаграмма исполняемых задач нескольких параллельно выполняемых операций с учетом потенциальных конфликтов «интересов» по ресурсам (например, по месту локализации АМР)

В определенной мере можно говорить о применении программно-информационных средств, подобных тем, которые используются в деятельности авиадиспетчеров, что, например, нашло обоснование в публикациях [27-29].

Заключение

В статье рассмотрены особенности взаимодействия ЧО с гетерогенной группой АМР в аспектах обеспечения безопасности их совместного функционирования с использованием концепции интеллектуального пространства. В качестве примера такой группы рассмотрено одно из возможных конструктивных решений по применению группы АМР на лунной поверхности и мониторингу их перемещений космонавтом, находящимся на удалении в лунной базе. При принятии решений по предотвращению коллизий АМР в качестве основного принимается антропоцентрический принцип распределения функций, когда текущие задачи, возложенные на отдельные АМР, могут быть скорректированы ЧО в отношении приоритетов их исполнения, инициирования, завершения и других видов управляющих команд ЧО. Для обеспечения скоординированных действий каждая задача АМР в составе полетной операции описывается онтологией, которая формализует основное предназначение и возможности АМР по сбору информации и информационному обмену, требования к ресурсам и ограничения по соображениям безопасности. При оценке текущей ситуации космонавтом данный подход позволяет использовать следующий объем знаний: 1) имеющиеся у ЧО варианты принятия решений по инициированию, приостановке, возобновлению, завершению задач в контексте понимания целей миссии и важности отдельных операций и задач в их составе; 2) условия реализации конкретной задачи в составе полетной операции, ресурсные ограничения в конкретные временные сроки; 3) пара-

метры движения и локализацию всех членов группы АМР в конкретные временные сроки; 4) синтезированное визуальное представление виртуального окружения с разметкой зон потенциального конфликта интересов на основе совокупности полученных данных об окружающей среде и текущей ситуации.

На этой технической основе в работе предлагается реализация мониторинга в рамках единой для всех АМР онтологии «управления на уровне задач» (включая этапы постановки, инициирования и завершения задач отдельных АМР) с учетом приоритетности полетных операций и технологических особенностей их реализации, что должно охватываться компетенциями ЧО, принимающего решения.

В рамках предлагаемого подхода при отображении виртуального окружения представляется целесообразным использовать визуальный образ ситуации, соответствующий ментальным представлениям человека о том, как должны взаимодействовать АМР, чтобы их активность отвечала принципам безаварийности и предотвращения коллизий.

В статье представлена онтологическая модель для организации знаний в области совместного решения задач человеком и роботами в экстремальной внешней среде. Для того, чтобы обеспечить совместное функционирование интеллектуальных мобильных роботов различного назначения, ответственность за подготовку и принятие решений о безаварийном управлении возлагается на человека с учетом его когнитивных возможностей ориентировки в слабо структурированной среде, образного мышления и способности к предвидению, а также специальной подготовке к выполнению диспетчерских функций.

Работа проводилась при поддержке гранта РФФИ 18-37-00306.

ЛИТЕРАТУРА

1. Левашова Т.В. Методология управления ресурсами интеллектуального пространства // НАУЧНЫЙ ВЕСТНИК НГТУ. Вып. № 1 (58) Январь - Март 2015. С.171-182. DOI: http://dx.doi.org/10.17212/1814-1196-2015-1-171-182
2. Lee, Joo-Ho & Hashimoto, Hideki. (2002). Intelligent Space – concept and contents. // Advanced Robotics. №16. pp. 265-280. DOI: 10.1163/156855302760121936.
3. Гаврилов А.В. Искусственный Домовой // Искусственный интеллект и принятие решений. Вып. 2. 2012, С. 77-89.
4. *Кашевник А.М., Пономарев А.В., Савосин С.В.* Управление гибридными системами на основе технологии интеллектуальных пространств // Труды СПИИРАН. 2014. Вып. 35. С. 212-226.
5. Al-Muhtadi J. Augmenting Smart Spaces with Intelligent Robots // 2007 2nd International Conference on Pervasive Computing and Applications. Birmingham, 2007, pp.505-510. DOI: 10.1109/ICPCA.2007.4365496
6. Carreira P., Resendes S., Santos A. Towards automatic conflict detection in home and building automation systems // Pervasive and Mobile Computing, vol. 12, 2014, pp. 37-57.
7. Кашевник А.М. Подход к обеспечению семантической интероперабельности мобильных роботов при формировании коалиций. // Информационные технологии и вычислительные системы. 2017. Т. 1. С. 90-100.
8. Михайлов С.А., Кашевник А.М. Организация интеллектуальных пространств на основе платформы Smart-M3 с использованием устройств на базе операционной системы DD-WRT // Труды СПИИРАН. 2017. № 3(52). С. 180-203.
9. Петров М.В., Кашевник А.М. Онтолого-ориентированный подход к непрямому взаимодействию пользователей и роботов для совместного решения задач. // Научный вестник Новосибирского государственного технического университета. 2017. № 1(66). С. 133-146.
10. Юсупов Р.М., Ронжин А.Л. От умных приборов к интеллектуальному пространству // Вестник Российской Академии Наук. 2010. Том 80. Вып. 1. С. 45–51.
11. Ронжин Ал.Л., Карпов А.А. Сравнительный анализ функциональности прототипов интеллектуальных пространств // Труды СПИИРАН. 2013. Вып. 1(24). С. 277-290.
12. Космонавтика XXI века / Под ред. Академика РАН Чертока Б.Е. – М.: Изд-во «РТСофт», 2010. 864 с.
13. Луна – шаг к технологиям освоения Солнечной системы / Под научн. ред. В.П. Легостаева, В.А. Лопоты. – М.: РКК «Энергия», 2011. 584 С.
14. Bartlett P., Wettergreen D., Whittaker W.L. Design of the Scarab Rover for Mobility and Drilling in the Lunar Cold Traps // Proceedings of the 9th International Symposium on Artificial Intelligence, Robotics and Automation in Space (iSAIRAS). 2008. URL: http://repository.cmu.edu/robotics/1104/ (accessed: 24.02.2015). (дата обращения 02.04.2018).
15. Bodkin, D.K., Escalera, P., Bocam, K.J. A human lunar surface base and infrastructure solution // AIAA SPACE Forum. September 2006. doi: 10.2514/6.2006-7336
16. Крючков Б.И., Усов В.М., Ярополов В.И., Сосюрка Ю.Б., Троицкий С.С., Долгов П.П. Об особенностях профессиональной деятельности космонавтов при осуществлении лунных миссий. // Пилотируемые полеты в космос № 2(19). 2016. С. 35-57.
17. Васильев А.В., Кондратьев А.С., Градовцев А.А., Даляев И.Ю. Исследование и разработка проектного облика мобильной робототехнической системы для проведения геологической разведки на поверхности Луны // Труды СПИИРАН. 2016. Вып. 45. С. 141- 156.

18. Васильев А.В. Мобильные мини-роботы разведки: текущее состояние, характерные черты и общие тенденции развития // Известия ЮФУ. Технические науки. 2010. № 3 (104). С. 119-124.
19. Карпов А.А., Крючков Б.И., Ронжин А.Л., Усов В.М. Проектирование взаимодействия человек-робот в составе единой команды космонавтов и автономных мобильных роботов на поверхности Луны. // Экстремальная робототехника. 2016. Т.1, № 1. С.71-81.
20. Ющенко А.С. Диалоговое управление роботами с использованием нечетких моделей // Интегрированные модели и мягкие вычисления в искусственном интеллекте: Сб. тр. V Междунар. науч.-практич. конф. (Коломна, 28-30 мая 2009 г.). М.: Физматлит, 2009. Т.1. С. 97-108.
21. Михайлов Б.Б. , Назарова А.В., Ющенко А.С. Автономные мобильные роботы - навигация и управление // Известия ЮФУ. Технические науки №2 (175), 2016. С.48-67.
22. Пшихопов В.Х., Соловьев В.В., Титов А.Е., Финаев В.И., Шаповалов И.О. Групповое управление подвижными объектами в неопределенных средах. – М.: Физматлит, 2015. 305 с
23. Тимофеев А.В. Адаптивные робототехнические комплексы. – Санкт-Петербург: Машиностроение, 1988. 332 с.
24. Герасимов В.Н., Михайлов Б.Б. Решение задачи управления движением мобильного робота при наличии динамических препятствий // Вестник МГТУ им. Н.Э. Баумана. Приборостроение. 2012. С.83-92.
25. Юревич Е.И. О проблеме группового управления роботами // Мехатроника, автоматизация, управление. 2004. № 2. С. 9-13.
26. Карпов А.А., Крючков Б.И., Ронжин А.Л., Усов В.М. Проектирование взаимодействия человек-робот в составе единой команды космонавтов и автономных мобильных роботов на поверхности Луны. // Экстремальная робототехника. 2016. Т. 1, № 1. С. 71–81.
27. Ворона А.А., Сыркин Л.Д., Крючков Б.И., Усов В.М. Визуальное представление космонавту перемещений группы автономных мобильных роботов на поверхности Луны для предотвращения их коллизий. // Пилотируемые полеты в космос. 2016 №3 (20). С.41-57.
28. Михайлюк М.В., Крючков Б.И., Усов В.М. Виды интерфейса для дистанционного взаимодействия космонавтов с автономными мобильными роботами при внекорабельной деятельности на лунной поверхности // Пилотируемые полеты в космос. 2017. №4(22). С.41-53.
29. Kryuchkov B.I., Usov V.M., Chertopolokhov V.A., Ronzhin A.L., and Karpov A.A. Simulation of the «Cosmonaut-Robot» system interaction on the lunar surface based on methods of machine vision and computer graphics //Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W4, 129-133, doi: 10.5194/isprs-archives-XLII-2-W4-129-2017, 2017.

B.I. Kryuchkov, V.M. Usov, D.V. Ivanko (Federal State Budgetary Organization “Gagarin Research & Test Cosmonaut Training Center”; Federal State Institution of Science “Saint Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences” (SPIIRAS))

Prospects for the use of smart spaces for information support of a human - operator at remote monitoring a group of mobile robots on the lunar surface

The article deals the application of the concept of intelligent space to the task management of a group of Autonomous mobile robots (AMR) on the lunar surface from the point of view of the human factor is considered. The anthropocentric principle of distribution of functions in the human-machine system "a cosmonaut – group AMR" is proposed for consideration in the planned extravehicular activity (EVA) on the lunar surface. In this case, in the process of monitoring the current situation by the cosmonaut, it is necessary to ensure his situational awareness and provide him with the authority to activate, suspend, stop certain tasks of the AMR (in consistently executed sequences according to the structure of flight operations), based on the main criterion – the accident-free functioning of the AMR group. To make a decision on possible conflicts of EVA participants and / or conflicts of interest at the level of accounting and resource planning, it is necessary to build an interactive environment, that provides modeling and visualization of a large set of data on the current situation obtained from active intelligent agents, in this case - from the AMR group, and on this basis to perform a synthesis of the virtual environment and operational display of interacting virtual agents. Among the main distinctive features, the intellectual environment to support management decision-making requires the development of ontological-oriented design tools, which significantly takes into account the role of the human factor in ensuring the safety of the joint functioning of a human and AMRs.

РАЗРАБОТКА БОРТОВОГО ВИРТУАЛЬНОГО СИМУЛЯТОРА ДЛЯ ФОРМИРОВАНИЯ И ПОДДЕРЖАНИЯ ГОТОВНОСТИ ЭКИПАЖА НА ЛУННОЙ СТАНЦИИ К АВАРИЙНОМУ РЕАГИРОВАНИЮ НА ПОТЕНЦИАЛЬНО ОПАСНЫЕ СИТУАЦИИ

М.В. МИХАЙЛЮК
ФГУ «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук» 117218, Москва, Нахимовский просп., 36, к.1. mix@niisi.ras.ru

Б.И. КРЮЧКОВ, В.М. УСОВ

ФГБУ «Научно-исследовательский испытательный центр подготовки космонавтов имени Ю.А. Гагарина», 141160, Россия, Звездный городок Моск. обл., b.kyuchkov@gctc.ru; v.usov@gctc.ru

Аннотация

Для повышения безопасности выполнения перспективных миссий освоения Луны путем формирования и поддержания готовности космонавтов к экстренному реагированию на потенциально опасные ситуации, а также для оказания им информационной поддержки при автономной работе на стационарной многомодульной лунной базе (СМЛБ), предлагается создание бортового виртуального симулятора (БВС) нештатных ситуаций (НшС), моделирующего опасные события, возникающих как следствие самих НшС, так и действий экипажа. Поскольку речь идет о предстоящих лунных миссиях, симулятор носит исследовательских характер и предназначается для моделирования условий принятия космонавтом тактических решений по выбору рациональных способов действий, направленных на устранение аварийной ситуации, сохранение живучести герметически замкнутого объекта и спасение экипажа при аварии.

В основе предлагаемой разработки лежит анализ применяемых в настоящее время моделей, прототипов и пространственно-ориентированных методов построения виртуальной среды.

Ключевые слова: стационарная многомодульная лунная база (СМЛБ); человек-оператор (ЧО); нештатная ситуация (НшС); бортовой виртуальный симулятор (БВС); мобильный робот-помощник

Введение

В настоящее время для повышения безопасности пилотируемых космических полетов предусмотрено проведение бортовых тренировок экипажа МКС для поддержание навыков управления пилотируемым космическим комплексом и его системами с учетом возможного воздействия факторов космического полета на организм человека и, в частности, на его сенсорную и двигательную сферы [1,2].

Важная составная часть бортовых тренировок – отработка действий экипажа при моделировании различных нештатных ситуаций (НшС), для своевременного обнаружения, идентификации, предупреждения аварийных ситуаций на орбитальной станции [3-6].

В перспективе при реализации планов полетов к Луне и проведении большого комплекса работ по ее освоению, включая построение стационарных многомодульных лунных баз (СМЛБ), можно ожидать применения аналогичного подхода для повышения готовности членов экипажа адекватно реагировать на НшС для ликвидации ее причин. В случае аварии необходимо принимать решение в автономных условиях, без поддержки с Земли, на спасение и эвакуацию экипажа из очага чрезвычайной ситуации (ЧС).

С учетом возможного использования в лунных миссиях роботов-помощников в качестве средств поддержки действий экипажа при ЧС необходимо дополнительно рассматривать вопросы применения аварийно-спасательных роботов.

Предлагаемый доклад посвящен проблемам создания виртуального окружения с помощью пространственно-ориентированных технологий и обеспечения ЧО информационной поддержкой, учитывающей достаточно сложные варианты реализации сценариев НшС, которые предполагается реализовать в БВС при моделировании поражающих факторов ЧС, наличного оборудования для защиты организма, спасения и эвакуации пострадавших в ЧС.

Концепция построения исследовательского симулятора НшС для условий СМЛБ

В историческом аспекте методология проведения бортовых тренировок и предназначенные для этих целей технические средства претерпели значительную эволюцию, связанную с улучшением характеристик вычислительной техники и ростом технических возможностей моделирования визуального представления рабочей среды орбитальной станции. При этом в первую очередь были получены результаты создания и применения бортового тренажера, включающего двухмерную модель интерьера космической станции, для имитации работы экипажа с бортовым оборудованием применительно к возникновению в полете ЧС (пожар, разгерметизация, загрязнение атмосферы) [3]. В последующем отчетливо выявилась тенденция применения моделей виртуального окружения с опорой на пространственно-ориентированные решения, ввиду наличия объективных трудностей для ЧО выполнения ориентировки в конструктивно сложном пространственном окружении герметически замкнутого объекта – многомодульной орбитальной станции. Именно это обстоятельство стимулировало использование новых подходов на основе 3D моделирования и 3D визуализации с высоким уровнем реалистичности не только в отношении конструкции гермообъекта и интерьера его отсеков и модулей, но и функционирования систем станции, сопровождающих развитие НшС. Возможно полное воспроизведение инструментальной информации, соответствующей имитируемой НшС, обеспечивался доступом ЧО к актуальной инструментальной информации с учетом расположения сенсоров, датчиков и приборных панелей, содержащих индикаторы и органы управления реального объекта. Как одно из перспективных направлений рассматривалось использование технологии «внешнего наблюдателя», благодаря которой в ходе обучения члены экипажа получают возможность изменять положение виртуальных камер и наблюдать происходящие события в разных отсеках, в разных проекциях. Как часть системы «погружения» ЧО в виртуальную среду применялась имитация перемещения виртуальных моделей человека в интерактивном режиме с виртуальным окружением. Такой методический прием направлен на индивидуализацию реагирования обучаемых, каждый из которых управляет своим виртуальным двойником (аватаром), когда предоставлена возможность работать на выделенном рабочем месте в составе тренажера с персональными рабочими местами обучаемых, космонавтов и специалистов Центра управления полетом (ЦУПа) путем создания моделирующего устройства в архитектуре «клиент-сервер» компьютерной сети [4-6]. Эти решения можно рассматривать как прототипы, но сегодня большее место в подготовке отводится групповым занятиям с имитацией участия специалистов ЦУП, в то время как для лунных проектов дополнительным значимым критерием готовности экипажа будет умение организовать свою деятельность в условиях НшС при отсутствии связи с Землей.

Прогрессу бортовых тренажеров способствовал общий тренд на использование технологий моделирования на наземных технических средствах подготовки космонавтов [7-10]. В ряде работ эти вопросы освещены и проанализированы достаточно подробно [5-6,11-12].

Обращает на себя внимание работа [6], в которой достаточно подробно изложены сведения о виртуальном 3D симуляторе, который используется для обучения наземного персонала бортовому реагированию на ЧС в центре управления Columbus (Германия). Для использования этой разработки, как прототипа для БВС на СМЛБ, особую ценность представляют представленные в ней основные сценарии аварийных ситуаций на МКС и соответствующая стратегия ожидаемого реагирования экипажа, а также некоторые приемы моделирования НшС и алгоритмы действий участников тренировок. В частности, отмечается необходимость выделения в особую группу ЧС медицинского характера, поскольку такие особенности требуют глубокого анализа мероприятий не только по экстренной эвакуации, но и проведения при необходимости мероприятий медицинской направленности. Для лунных экспедиций вопросы оказания неотложной медицинской помощи пострадавшим в ЧС и их эвакуации в безопасную зону будут рассматриваться в числе первоочередных [13-15]. Применительно к МКС экипаж после обнаружения на борту опасной ситуации создает, так называемую, «безопасную зону», в которой исключено воздействие патогенного фактора ЧС и имеется необходимое аварийно-спасательное оборудование. При этом у космонавтов есть переносной компьютер управления космической станцией, а также предусмотрен маршрут эвакуации для перехода в космический транспортный корабль «Союз-Т», выполняющий функции

спасательных кораблей космической станции. По усмотрению командира принимается решение о том, следует ли экипажу либо начать с процедур экстренного реагирования для локализации источника ЧС (остановить/ограничить распространение), либо немедленно покинуть станцию для возвращения на Землю на транспортном корабле. Применительно к лунным экспедициям роль конечного звена эвакуации будет, возможно, выполнять взлетно-посадочный лунный модуль (ВПЛП), а маршрут эвакуации будет состоять из начального этапа внутри СМБЛ к защитной «зоне безопасности» и завершающего этапа на лунной поверхности к ВПЛП. На этом этапе ожидается применение аварийно-спасательных роботов.

В результате ЧС на лунной базе можно ожидать развития неблагоприятных или даже опасных состояний, которые требуют экстренной эвакуации и оказания неотложной медицинской помощи. Предполагается, что помимо автономной системы жизнеобеспечения в зоне безопасности могут быть сконцентрированы все необходимые средства для купирования опасных состояний пострадавшего. В работе [16] предложена систематизация состояний, связанных с аварийными ситуациями в космических полетах, которые могут иметь место и на объектах лунной инфраструктуры. В [16] выделены наиболее опасные состояния космонавта, требующие экстренной диагностики и медицинской помощи, которые могут возникнуть при разгерметизации, нарушениях в работе систем жизнеобеспечения, попадании микрометеоритов и при радиационных поражениях. В их числе названы следующие:

1) Острая гипоксия. Причинами ее развития могут быть: разгерметизация жилых отсеков, повреждения скафандра, неполадки в клапанах систем жизнеобеспечения и др.
2) Взрывная декомпрессия. Она возникает в случаях падения давления в герметическом объёме при повреждении стенки космического аппарата крупным метеоритом или при повреждении скафандра.
3) Высотная декомпрессионная болезнь. Она возникает при разгерметизации жилых отсеков станции при достаточно быстром понижении барометрического давления и кратности перепада давления более 2,25.
4) Состояния, связанные с повышением и понижением температуры тела членов экипажа, могут иметь место при нарушении системы терморегуляции в космическом комплексе.
5) Ожоги вследствие термических, электрических и химических воздействий.

Следствием патогенного воздействия целого ряда факторов ЧС в пилотируемых полетах могут быть жизнеугрожающие состояния, и в их числе – остановка кровообращения и дыхания; травматический и ожоговый шок; интоксикации и др. Из этого следует, что в составе медицинского оборудования должны быть средства для проведения мероприятий неотложной медицинской помощи: восстановления кровообращения, искусственного дыхания, введения противошоковых препаратов и лекарств для обезболивания. До настоящего времени вопросы имитации нештатных медицинских ситуаций в полной мере не изучены, ввиду их особой сложности и того факта, что для орбитальных полетов предусмотрены достаточно надежные средства предотвращения потери работоспособности членов экипажа. Для перспективных БВС необходимо моделировать различные конфигурации таких «зон безопасности» на СМЛБ, а также, возможно, зону проведения неотложных медицинских мероприятий («медицинская зона» для эвакуации), в которой помимо аварийно-спасательного оборудования и защитного снаряжения будет храниться запас медикаментов, аппаратура для сердечно-легочной реанимации, детоксикации, проведения искусственного дыхания, срочных хирургических вмешательств и др. [17-19]. Кроме того, в [17,18] отмечается необходимость использования бортовых симуляторов для отработки медицинских манипуляций в условиях автономного полета, что также может входить в состав инструментария и оборудования зоны.

В свете перечисленных особенностей возникновения ЧС на СМЛБ одним из критически важных направлений моделирования должно быть построение маршрутов эвакуации при различных исходных предположениях о местонахождении членов экипажа, роботов – помощников, местах возникновения ЧС и их последствий, приводящих к трудностям транспортировки людей и грузов пр. [20]. Тактические задачи будут состоять в не только в попытках поиска путей скорейшей локализации последствий аварии, но и в оценке возможностей сокращения времени на покидание помещений, попадающих в зону риска, а также минимизации времени на перемещение пострадавших членов экипажа в «зону

безопасности». Здесь особую роль при транспортировке может сыграть применение роботов-помощников, выступающих в роли аварийно-спасательных роботов [13].

Эти предложения по моделированию эвакуации могут найти реализацию в форме виртуальных моделей роботов в виртуальной среде, в которой они функционируют [20-24].

Применительно к МКС порядок действий экипажа при пожаре, аварийной разгерметизации и появлении токсических примесей в атмосфере станции достаточно детально представлен в работе [6]. Обращает на себя внимание требование к экипажу уметь выполнять классификацию ожидаемого ЧС после аварийного оповещения, поскольку это во многом определяет дальнейший выбор поведения членов экипажа. По этим данным возникает предположение о необходимости моделирования таких угрожающих явлений, как распространение пожара в гермообъекте, падение парциального давления кислорода в отсеках при различной скорости утечки газа при аварийной разгерметизации, а также визуального представления на схеме помещений в виде 2D или 3D-картины пригодности рабочей среды для жизнедеятельности.

Эта методология может рассматриваться, как исходная, применительно к условиям пребывания экипажа в условиях СМЛБ с необходимыми коррекциями в отношении того, что эвакуация экипажа будет иметь много новых особенностей, определяемых уровнем развития лунной инфраструктуры [25-27]. Такая постановка задачи предполагает наличие в составе программного обеспечения в составе взаимосвязанных программных модулей:

– поддержания интерактивности взаимодействия ЧО с виртуальной средой с применением многомодальных человеко-машинных интерфейсов;

– моделирования и 3D визуализации виртуального окружения в реальном масштабе времени в разных системах координат в зависимости от информационных потребностей ЧО;

– моделирования управления виртуальным роботом;

– моделирования физических процессов, связанных с распространением поражающих факторов и информационной обстановки в реальном масштабе времени.

Применительно к тренировкам персонала различных потенциально опасных объектов с использованием виртуальных симуляторов такие решения можно найти в работах [6,11,28-30]. Например, в работах [29,30] описана реализация на базе технологий виртуальной и дополненной реальности модели, позволяющей правильно: а) оценивать риски возникновения аварийной ситуации; б) прогнозировать ее развитие; в) действовать в аварийных ситуациях.

Для обнаружения и ликвидации аварийных ситуаций на борту космического аппарата в работе [31] сформированы основные требования к автоматизированной системе поддержки принятия решений в аварийных ситуациях на МКС и наиболее важные параметры, характеризующие аварийную ситуацию разгерметизации. Такие подходы имеют большую ценность для разработки симуляторов и тренажеров, например, применительно к МКС – для формирования и восстановления навыков ситуационной поддержки принятия решений сменным руководителем полетов и специалистами главной оперативной группы управления (ГОГУ) ЦУП, работающего в автономном цикле управления, а также проектирования деятельности экипажа [32].

Тем самым, существующие методы и средства подготовки экипажа и специалистов обеспечения полета к аварийным ситуациям на МКС можно с полным основанием рассматривать как прототипы для построения БВС, но с той существенной разницей в постановке задачи, что полномочия принятия решений передаются экипажу, работающему в автономном режиме без поддержки с земли и с существенно более ограниченными возможностями консультации.

Необходимо отметить некоторые отличия БВС в сравнении с применяемыми на МКС бортовыми тренажерами, поскольку в данном случае на первый план выходят вопросы моделирования в реальном масштабе времени и реалистичной визуализации ситуационной обстановки при широком спектре начальных условий.

Общая схема построения БВС и его использования представлена на рис.1.

Дополнительно к представленным на этой схеме подсистемам необходимо найти место в имитации обстановки современных методов моделирования процессов распространения поражающего фактора и эвакуации в пожароопасных ситуациях. Необходимо отметить, что многие предлагаемые в этих источниках вычислительные модели вызывают трудности интерпретации результатов конечным пользователем – лицом, принимающим решения в

сложной обстановке. Это требует разработки графического интерфейса, который позволяет осуществлять визуализацию помещений, изменять их размеры и конфигурацию, а также получать динамическую картину распространения поражающего фактора с учетом возможных путей эвакуации.

Состав основных подсистем исследовательского виртуального симулятора

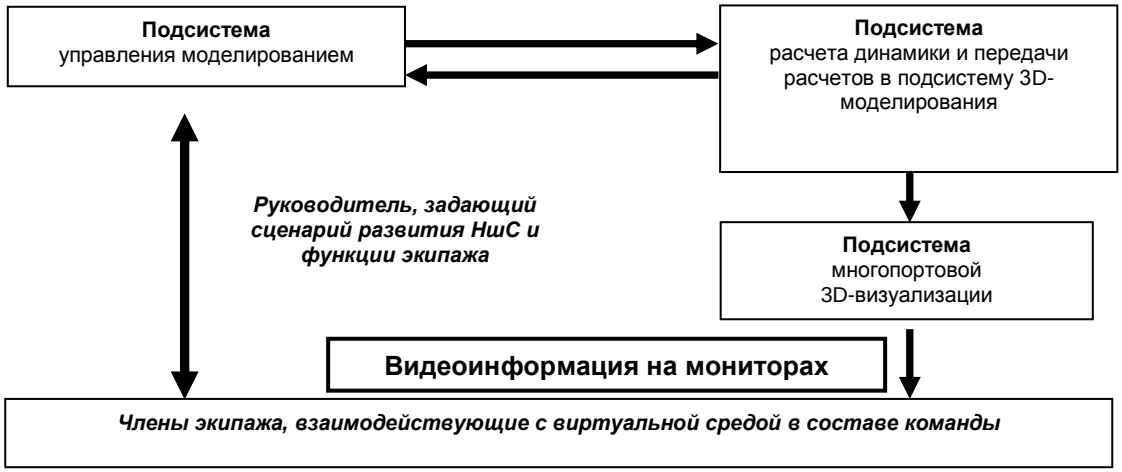


Рис.1. Функциональная схема построения и использования БВС.

Не менее важно ввести в состав симулятора модели разгерметизации. При разгерметизации основными параметрами, характеризующими возникновение и развитие аварийной ситуации, являются изменение давления в отсеках станции и допустимое время нахождения экипажа в разгерметизированном отсеке [31]. Согласно этой работе, в процессе моделирования такой ситуации должен учитываться набор наиболее важных характеристик: объем аварийного сегмента; темп падения давления в этом сегменте; текущие и прогнозные значения давления в каждом из отсеков станции; резервное (допустимое) время нахождения экипажа в отсеке. Сценарий действий членов экипажа регламентирует последовательность действий в зависимости от результатов измерения давления в соответствующих сегментах.

В представленной выше общей постановке назначения исследовательского симулятора для условий автономной работы экипажа можно обозначить контуры тех режимов, которые повышают готовность космонавтов к экстренному реагированию на НшС, и формируют у них оперативное мышление на базе предлагаемой схемы анализа тактической обстановки, что способствует более полному учету факторов риска при спасении и эвакуации людей их зоны воздействия опасных для жизни и здоровья факторов.

Первым уровнем изучения обстановки является работа ЧО с моделью космического комплекса (в данном случае – СМЛБ), представленной в 2D или 3D формате, с использованием виртуальных TV-камер. Дополнительно может использоваться многопортовый вывод информации на монитор на рабочем месте космонавта. Это позволяет ЧО методом постоянных подкреплений усвоить данные, относящиеся к топологии комплекса и принимать во внимание наличие переходов между модулями и секциями СМЛБ при передвижении в условиях заданных ограничений на время операции. Важно также показать локализацию шлюзов и возможность перекрытия переходов, изоляции отдельных отсеков и модулей. Этот режим является прологом к освоению разнообразных способов построения рациональных маршрутов быстрого доступа к месту возникновения отказов оборудования, очагам возгорания или, напротив, путей эвакуации при различном местоположении космонавтов и различной локализации очага опасности. Важный новый нюанс подготовки космонавта к экстренному реагированию – расчет и отображение на индикаторе расчетного времени передвижения по выбранному маршруту (исходя из известных скоростей перемещения космонавта, облаченного в различное спецснаряжение) и, соответственно, обратный отсчет резерва времени, когда речь идет о риске снижения работоспособности и/или критическом нарушении состояния.

Представляется, что новыми компонентами симулятора будут модели:

– распространения поражающих факторов в зависимости от конфигурации комплекса и доступности секторов, наличия нарушений целостности стенок и особенностей режимов вентиляции и регуляции правильного состава газов искусственной атмосферы;

– эвакуации в зависимости от строительных и технологических конструкций, а также динамически изменяемой конфигурации комплекса;

– хранимых знаний об уровне рисков, основанных на онтологиях предметной области поддержания обитаемости и обеспечения безопасности СМЛБ;

– активных интеллектуальных агентов, которые могут принять участие в проведении спасательных и эвакуационных мероприятий как внутри СМЛБ, так и при ВКД при "дальней" эвакуации пострадавших в лунный взлетно-посадочный модуль.;

– интерактивных виртуальных операторов (аватаров) в виртуальном окружении;

– человеко-машинных интерфейсов, позволяющих управлять на симуляторе визуальным отображением обстановки на основе многооконного многопортового способа индикации.

В настоящее время проводится разработка системы моделирования и визуализации для решения задачи построения симулятора НшС применительно к СМБЛ и объектам на лунной поверхности (лунный взлетно-посадочный модуль, мобильные автономные роботы, стационарно размещаемые объекты инфраструктуры).

Разработан дизайн ряда объектов, которые предполагается использовать при визуализации виртуального окружения (рис.2).

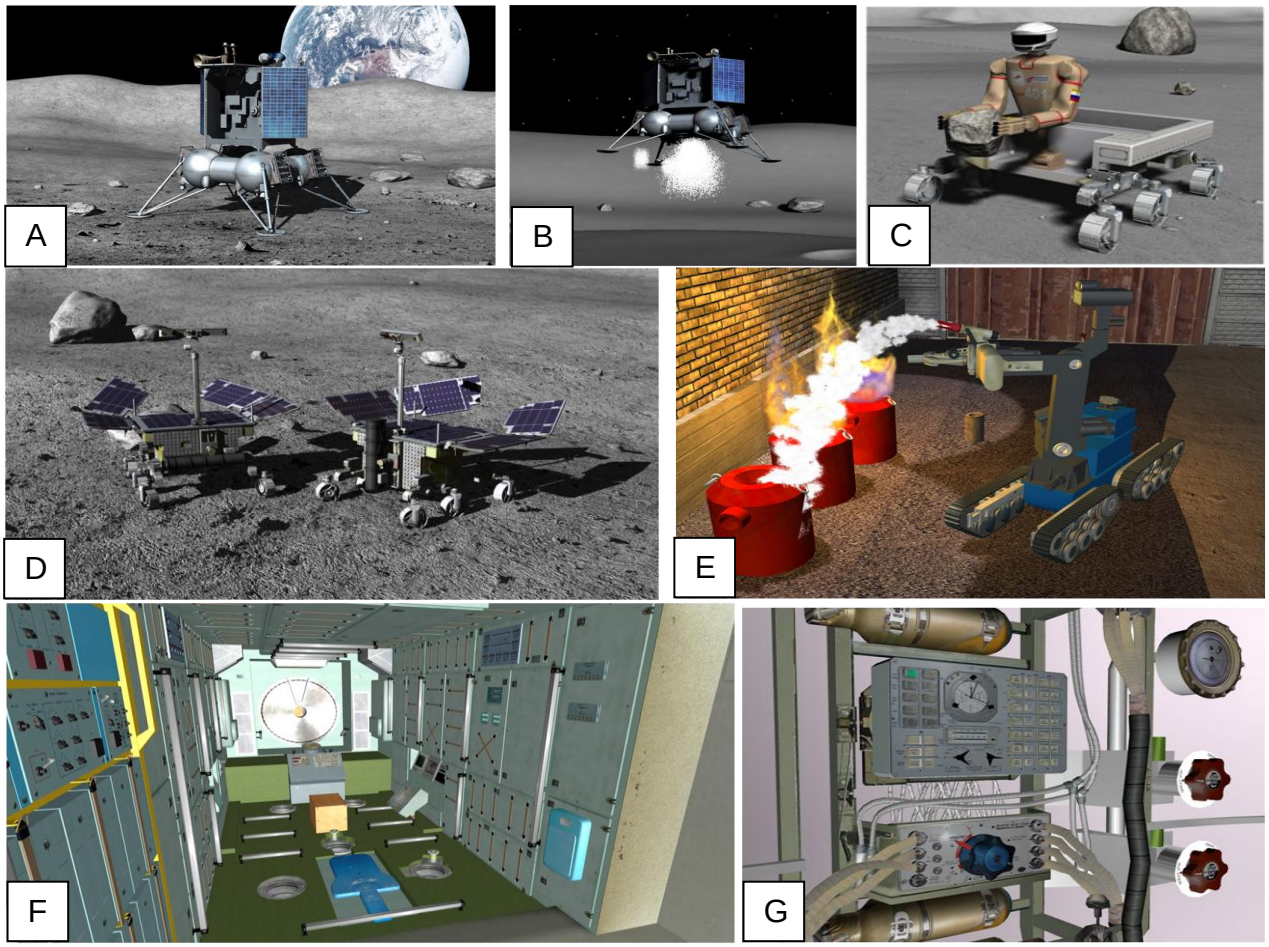


Рис.2. Дизайн имитируемых в БВС объектов для интерактивной визуальной среды.

Где: A, B – дизайн взлетно-посадочного лунного модуля, который может рассматриваться как объект для эвакуации экипажа при ЧС на СМЛБ;

C – дизайн мобильных автономных роботов для внекорабельной деятельности,

D – дизайн антропоморфного робота на колесной платформе для перемещения грузов и эвакуации пострадавших по лунной поверхности;

E – дизайн мобильного (пожарного) робота для ликвидации пожара (показаны очаги возгорания и их тушение пеной);

F – интерьер одного из технологических отсеков;

G – дизайн пульта управления бортовым оборудованием.

Заключение

Человеческий фактор является основным риск-фактором при моделировании событий, направленных на повышение безопасности эксплуатации сложных человеко-машинных систем. Сегодня разработка тактических симуляторов исследовательского характера, моделирующих космические операции в целях повышения безопасности пилотируемых полетов – важное прикладное направление исследований надежности профессиональной деятельности космонавтов и составная часть современной системы их подготовки к космическим полетам.

В соответствие со сценариями прогнозируемых аварийных ситуаций, связанных с высоким уровнем риска для жизни и здоровья космонавтов при их пребывании на лунной базе, в фокусе внимания предлагаемого методического подхода и возможных технических решений находятся приемы формирования разнообразных тактических задач и поиска их решения в «игровом режиме» взаимодействия космонавтов с виртуальными моделями активных агентов в виртуальном окружении. Такие принципы в настоящее время практикуются для подготовки бригад для ликвидации последствий ЧС. В качестве прототипов для БВС НшС на лунной базе в данной работе рассматриваются бортовые тренажеры МКС аналогичного предназначения, но с существенными отличиями, как методического характера, так и по условиям эксплуатации. Первой особенностью рассматриваемого БВС является предположение о необходимости выбора тактики действий в автономных условиях пребывания на лунной базе, без оперативной поддержки специалистов Центра управления полетом. Вторая особенность – моделирование не только действий активных агентов («аватаров»– компьютерных моделей людей и роботов) в виртуальном 3D-окружении, но и воспроизведение в визуальном представлении физических явлений, сопровождающих пожар, загрязнение искусственной атмосферы в гермообъекте и др. Третья особенность – предположение о необходимости моделирования событий не только на лунной базе, но и на лунной поверхности при ВКД, то есть во внешней среде, когда возникает необходимость эвакуации экипажа из очага ЧС во взлетно-посадочный лунный модуль с использованием аварийно-спасательных роботов. В этом аспекте предлагается использовать виртуальные модели мобильных роботов-помощников, способных оказать помощь экипажу как внутри гермообъекта, так и на лунной поверхности с учетом вариантов построения маршрутов эвакуации. В статье рассмотрены технологии построения тактического симулятора с использованием концепции построения и визуализации виртуального окружения модели СМЛБ с внесением элементов моделирования неблагоприятных событий на основе технологий дополненной и индуцированной реальности. Именно это обстоятельство учитывается в первую очередь при формулировке требований к применению программно-информационных средств на основе 3D визуализации для имитации НшС и отработке приемов реагирования экипажа на опасные события, в том числе, требующих учета медицинских рисков. Практическая разработка программно-аппаратных средств БВС и дизайн объектов визуального отображения составляют предмет дальнейших исследований.

ЛИТЕРАТУРА

1. **Сальницкий В.П., Мясников В.И., Бобров А.Ф., Шевченко Л.Г., Дудукин А.В.** Исследование надежности деятельности космонавта на различных этапах длительного космического полета (эксперимент «Пилот») // Орбитальная станция «Мир». Космическая биология и медицина. – Т. 2. Медико-биологические эксперименты. – М.: ГНЦ РФ–ИМБП РАН, 2002. С. 285-300.
2. **Сальницкий В.П., Бронников С.В., Городецкий И.Г.** Психодиагностический комплекс-тренажер для оценки и прогнозирования надежности профессиональной деятельности космонавта // Приборы. 2008. № 4. С. 23-28.
3. **Бронников С.В., Рожкова И.А.** Тренажер аварийных ситуаций // Доклад на XLIV научных чтениях памяти К.Э.Циолковского, Калуга 15-17 сентября, 2009. С.100-101.
4. **Бронников С.В., Шевченко Л.Г., Рожкова И.А**. Развитие технического и методического обеспечения бортовых тренировок экипажей по действиям при разгерметизации космической станции // Научные чтения памяти К.Э.Циолковского, 2004 г., Секция «К.Э. Циолковский и проблемы профессиональной деятельности космонавтов», Россия, г. Калуга, Государственный музей истории космонавтики им. К.Э. Циолковского. //

Электронный ресурс: URL: http://readings.gmik.ru/lecture/2004-RAZVITIE-TEHNICHESKOGO-I-METODICHESKOGO-OBESPECHENIYA-BORTOVIH-TRENIROVOK-EKIPAZHEY-PO-DEYSTVIYAM-PRI-RAZGERMETIZATSII-KOSMICHESKOY-STANTSII Доступ свободный, 12.03.2018.

5. **Бронников С.В., Маликов А.Е., Рожкова И.А.** Подготовка космонавтов к действиям в аварийных ситуациях на космической станции // Пилотируемые полеты в космос. №2(27). 2018.С.35-45.
6. **Uhlig, Th. Uhlig Th., Roshani Fr.-C., Amodio C., Rovera A., Zekusic N., Helmholz H., Fairchild M.** (2016) ISS emergency scenarios and a virtual training simulator for Flight Controllers. // Acta Astronautica (128), Seiten 513-520. Elsevier. DOI: 10.1016/j.actaastro.2016.08.001.
7. **Karpov A.A., Mikhaylyuk M.V., Usov V.M., Kryuchkov B.I.** Virtual prototyping of Human-Machine Interaction for remote control of space autonomous manipulation robots based on augmented reality technology // IAC2016, Paper number IAC-16,B3,6-A5.3,7,x34571 URL: https://iafastro.directory/iac/archive/browse/IAC-16/B3/6-A5.3/34571/ Date: 21.11.3016.
8. **Крючков Б.И., Михайлюк М.В., Усов В.М.** Технологии моделирования для эргономического проектирования системы «космонавт – манипуляционный робот – рабочая среда» // Материалы конференции «Управление в морских и аэрокосмических системах» (УМАС-2014). СПб.: ОАО «Концерн «ЦНИИ «Электроприбор», 2014. (871 с.) С.367-377.
9. **Крючков Б.И., Усов В.М.** Создание моделей виртуальной реальности, как способ обучения космонавтов взаимодействию с роботом - помощником и как условие определения потенциальных областей его полезного применения // Робототехника и техническая кибернетика. 2013. №1(1). С.46-52.
10. **Бурдин Б.В., Михайлюк М.В., Сохин И.Г., Торгашев М.А.** Использование виртуальных 3D-моделей для экспериментальной отработки бортовых полетных операций, выполняемых с помощью антропоморфных роботов // Робототехника и техническая кибернетика. 2013. №1(1). С.42–46.
11. **Di-Ping Yuan, Ming-Jun Wu, Li-Biao Pan.** (2014) Design and Implementation of Virtual Training Scenario Editing System of Emergency Disposal // 7th International Conference on Intelligent Computation Technology and Automation 25-26 Oct. 2014 pp.174-177. DOI: 10.1109/ICICTA.2014.49
12. **Kuch T., Sabath D.** The Columbus-CC - Operating the European laboratory at ISS. //Acta Astronautica, Vol. 63, 2008. pp. 204-212.
13. **Поляков А.В., Алтунин А.А., Крючков Б.И., Мотиенко А.И., Ронжин А.Л., Усов В.М.** Использование роботов-спасателей при развитии медицинских нештатных ситуаций во время внекорабельной деятельности на поверхности Луны // Авиакосмическая и экологическая медицина. 2018. Т.52. № 2. С.34-41. DOI: 10.21687/0233-528X-2018-52-2-34-41.
14. **Крючков Б.И., Усов В.М., Ярополов В.И., Сосюрка Ю.Б., Троицкий С.С., Долгов П.П.** Об особенностях профессиональной деятельности космонавтов при осуществлении лунных миссий / // Пилотируемые полеты в космос. 2016 №2 (19). С.35-57.
15. **Карпов А.А., Крючков Б.И., Ронжин А.Л., Усов В.М.** Проектирование взаимодействия человек–робот в составе единой команды космонавтов и автономных мобильных роботов на поверхности Луны / // Экстремальная робототехника. // Труды Международной научно-технической конференции. СПб., 24-26 ноября 2016 г. – Санкт-Петербург: ООО «АП4Принт», 2016. (480 с.) С. 76-80.
16. **Егоров А.Д.** Теория и методологии медицинского контроля в длительных космических полетах (Актовая речь). Москва. ГНЦ РФ – ИМБП РАН. Октябрь-2001 / Электронный ресурс. URL: http://www.imbp.ru/webpages/win1251/science/uchsov/docl/2001/egorov_actsp.html (Доступ свободный, дата 2018.06.05).
17. **Ушаков И.Б., Поляков А.В., Усов В.М., Бояринцев В.В.** Перспективы использования на пилотируемых космических аппаратах симуляционных технологий для практического освоения и поддержания в полете навыков выполнения космонавтами мероприятий неотложной медицинской помощи // Авиакосмическая и экологическая медицина. 2015. Т.49. №2. С.50-54.
18. **Ушаков И.Б., Поляков А.В., Карпов А.А., Усов В.М.** Медицинская робототехника как новый этап развития бортовых тренажеров и биотехнических систем на орбитальной станции // Труды международной научно-технической конференции «Экстремальная робототехника». Санкт-Петербург: Изд-во «Политехника-сервис», 2015. С. 43-53.
19. **Поляков А.В., Грязнов Н.А., Сенчик К.Ю., Усов В.М., Мотиенко А.И.** Ассистивные возможности роботизированных систем для сердечно-легочной реанимации в условиях лунной базы // Авиакосмическая и экологическая медицина. 2018. Т. 52. № 3. С. 13-27. DOI: 10.21687/0233-528X-2018-52-3-13-27.
20. **Крючков Б.И., Карпов А.А., Усов В.М.** Перспективные подходы к применению сервисных роботов в области пилотируемой космонавтики. // Труды СПИИРАН. 2014. Вып. 32. С.125-151.
21. **Торгашев М.А**. Моделирование огня и средств пожаротушения для задач распределенной симуляции робототехнических средств в режиме реального времени // САИТ-2013: тр. V Междунар. конф. Красноярск, 2013. Т.1. С. 293-298.
22. **Ворона А.А. , Сыркин Л.Д. , Крючков Б.И. , Усов В.М.** Визуальное представление космонавту перемещений группы автономных мобильных роботов на поверхности Луны для предотвращения их коллизий. // Пилотируемые полеты в космос. 2016. №3 (20). С.41-57.
23. **Сыркин Л.Д., Усов В.М., Крючков Б.И., Ворона А.А.** Эргономические аспекты синтеза систем отображения внешней обстановки оператору при дистанционном управлении автономными мобильными роботами // Кибернетика и программирование, 2016. № 3. С.76-92.
24. **Мальцев А.В.** Методы высокореалистичной визуализации для космических тренажерных комплексов // XXXVII Гагаринские чтения. Научные труды Международной молодежной научной конференции. М., 2011. С. 84-86.

25. **Градовцев А.А., Кондратьев А.С., Лопота А.В.** Средства робототехнического обеспечения перспективной космической инфраструктуры // Научно-технические ведомости СПбГПУ. 2013. № 1. С. 111-118.
26. **Легостаев В.П., Лопота В.А. и др.** Луна – шаг к технологиям освоения Солнечной системы / Под науч. ред. Легостаева В.П., Лопоты В.А. – М.: РКК «Энергия», 2011. 550 с.
27. **Брюханов Н.А., Легостаев В.П., Лобыкин А.А., Лопота В.А., Сизенцев Г.А., Синявский В.В., Сотников Б.И., Филиппов И.М., Шевченко В.В.** Использование ресурсов Луны для исследования и освоения Солнечной системы в XXI веке // Космическая техника и технологии. 2014. № 1(4). С. 3–14.
28. **Хафизов А.М., Кошелев Н.А., Гумеров Д.А., Крышко К.А., Сидоров Д.А.** Разработка виртуального тренажера – имитатора возникновения аварийной ситуации – снижения расхода нагреваемого продукта в трубчатой нагревательной печи // Фундаментальные исследования. 2016. №10-3. С.576-580. // Электронный ресурс: URL: http://www.fundamental-research.ru/ru/article/view?id=40898 (доступ свободный, дата обращения: 29.05.2018).
29. **Горбунова А.А.** Безопасное моделирование чрезвычайных ситуаций в реальной среде посредством дополненной реальности// Виртуальная и дополненная реальность-2016: состояние и перспективы / Сборник научно-методических материалов, тезисов и статей конференции. Под общей редакцией д.т.н., проф. Д.И. Попова. - М.: Изд-во ГПБОУ МГОК, 2016. (386 с.) С.110-113. / Электронный ресурс: URL: http://arconf.hi-edu.ru/ru/sbornik_mgok_AR_VR_conf.pdf (доступ свободный, дата 2018.06.05.)
30. **Горбунов А.Л., Елисеев Б.П., Нечаев Е.Е.** Инновационный подход к созданию авиадиспетчерских тренажеров на базе технологий комбинированной реальности: необходимость и возможность // Научный вестник МГТУ ГА. 2010. №161. С.5-14.
31. **Матюшин М.М., Потоцкий С.И. Лахин О.И.** Автоматизированная система поддержки принятия решений в аварийных ситуациях // Программные продукты и системы. № 3. 2013. С.61-68.
32. **Юрыгина Ю.С., Коршиков Д.Н., Носкова А.И.** Адаптивный тренажер для формирования и восстановления навыков ситуационной поддержки принятия решений сменным руководителем полетов и специалистами ГОГУ // Тезисы докладов XX научно-технической конференции молодых ученых и специалистов, ОАО «Ракетно-космическая корпорация «Энергия» имени С.П. Королёва», 10 – 14 ноября 2014 года. – С. 629-631.

M. V. Mikhaylyuk, Federal state University "Federal research center research Institute of system research Russian Academy of Sciences" 117218, Moscow, Nakhimovsky Ave., 36, K. 1.
+7 903 298-2146 mix@niisi.ras.ru
B. I. Kryuchkov, V. M. Usov, Federal state Unit " Gagarin Scientific Research &Test Cosmonaut Training Center", 141160, Russia, Star city, Moscow region,
+7 917 550-4807 Email: b.kyuchkov@gctc.ru ; +7 910 459-4902 Email: v.usov@gctc.ru

The development of on-board virtual simulator to build and maintain readiness of the crew to the emergency response on a potentially dangerous situation on the lunar station

Annotation

To improve the safety of the promising missions of the Moon by forming and maintaining the readiness of cosmonauts for an emergency response to potentially dangerous situations, as well as for providing them with information support for autonomous operations on a stationary multi-module lunar base (SMMLB), in the article proposed to create the virtual simulator (VS) of emergency situations (ES), for enable modeling dangerous events, arising as a consequence of the ES and on account of the crew's activity. Since we are talking about the upcoming lunar missions, the simulator is of a research nature and is designed to simulate the conditions of the cosmonaut's tactical decisions on the choice of rational methods of action aimed at eliminating the ES, maintaining the survivability of a hermetically closed object and saving the crew in an accident. The proposed development is based on the analysis of currently used models, prototypes and spatially-oriented methods of building a virtual environment.

Keywords: stationary multimodule lunar base (SMMLB); Human Operator (HO); emergency situation (ES); onboard virtual simulator (VS); mobile robot assistant of a crew.

Секция 8
ПРОБЛЕМЫ ИНЖЕНЕРНОГО ОБРАЗОВАНИЯ В ОБЛАСТИ УПРАВЛЕНИЯ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ: СОВРЕМЕННОЕ СОДЕРЖАНИЕ И ТЕХНОЛОГИИ ОБУЧЕНИЯ

ПРОФЕССИОНАЛЬНАЯ ПОДГОТОВКА ИЛИ ОБРАЗОВАНИЕ?

А. Г. Степанов
(*Санкт-Петербургский государственный университет аэрокосмического приборостроения (ГУАП). 190000, С.-Петербург, ул. Большая Морская, д.67, лит. А., georgich_spb@mail.ru*)

Аннотация
Рассматривается проблема возможной потери фундаментальности образования при внедрении в практику новых федеральных государственных стандартов высшего образования, основанных на профессиональных стандартах министерства труда и социальной защиты

Ключевые слова: ФГОС, профессиональный стандарт, фундаментальность образования.

Введение

Основной задачей высшей школы было и, надеемся, будет именно образование, относящееся к фундаментальным правам человека, сохраняющее и развивающее духовные и материальные достижения человечества и представляющее собой целенаправленный процесс обучения и воспитания в интересах личности, общества и государства. Естественная потребность в получении дальнейшего образования заставляет выпускников школ продолжать обучение в университетах. Вполне логично, что некоторые из них (но не обязательно все) далее будут реализовывать свои способности на практике в том числе в процессе трудовой деятельности. С течением времени эти люди станут профессионалами, которые будут развивать избранные ими научные направления и получать удовлетворение от своей производственной деятельности.

С другой стороны, профессиональная деятельность формулируют к работнику, занимающему определенное место в производственной иерархии, требования в виде утвержденных министерством труда и социальной защиты профессиональных стандартов. Они регламентируют необходимые трудовые функции и детально проработанные соответствующие им трудовые действия, а также необходимые умения и знания. Отметим явное противоречие между двумя возможными целями обучения в вузе: индивидуальной – получение образования, и общественно значимой – подготовка квалифицированных кадров. Это противоречие в конечном итоге сводится к противоречию между необходимостью обеспечения фундаментальной и специализированной подготовкой будущих выпускников (бакалавров, магистров).

Предлагаемый доклад посвящен анализу отличий подходов действующего в настоящее время и вводимого в следующем учебном году варианта стандарта инженерной подготовки по направлению Прикладная информатика, которые сводятся, в конечном итоге, либо к созданию системы высшего инженерного образования, либо к профессиональной подготовке выпускников для конкретных рабочих мест.

Два варианта построения системы обучения

Действующие сегодня ориентированные на компетентностный подход стандарты образования привели к существенному изменению технологии работы высшей школы. Реализуя ее вузы получили возможность самостоятельно формировать учебный план и, как следствие, определять содержание обучения. Существенные изменения коснулись всех видов аттестаций студентов. Привычные экзамены и зачеты (контрольно-измерительные материалы) стали вспомогательным средством контроля имеющихся знаний, умений и навыков. На первый план выдвинулись так называемые оценочные средства, которые используются для контроля за результатами формирования регламентированных стандартом компетенций выпускников. Специфика такого

контроля заключается в том, что реализовать его можно только оценивая результаты творческой работы студента. К ним относятся, в первую очередь, курсовые работы, отчеты по практикам, выпускные квалификационные работы, а также некоторые другие результаты самостоятельной деятельности (рефераты, эссе, доклады, статьи, экспонаты выставок и т.п.) [1]. Учитывая все еще существующий научный потенциал вузов, действующие сегодня стандарты позволили организовать инженерную подготовку с учетом конкретных возможностей университетов сохранив фундаментальность обучения.

Следуя своему негласному правилу один раз в пять лет менять стандарты высшего образования, в конце прошлого года тогда еще министерство образования и науки утвердило и ввело в действие новый стандарт подготовки по направлениям 09.03.03 и 09.04.03 Прикладная информатика уровней бакалавриата и магистратуры. В отличие от предыдущих, этот стандарт регламентируют области профессиональной деятельности выпускников высшей школы основывающиеся, в свою очередь, на профессиональных стандартах министерства труда и социальной защиты. К таким областям применительно для выпускников по направлению Прикладная информатика отнесены 06 Связь, информационные и коммуникационные технологии и 40 Сквозные виды профессиональной деятельности в промышленности. Фактически речь идет о внедрении в практику высшей школы системы профессиональной подготовки в конкретных областях как совокупности специальных знаний, умений, навыков и компетенций, позволяющих выполнять работу в конкретной достаточно узкой области профессиональной деятельности.

Отметим, что реальный выпуск специалистов по направлению Прикладная информатика существенно превышает потребности перечисленных в стандарте областей профессиональной деятельности. Как следствие, выпускники вузов вынуждены будут искать себе другие рабочие места, которые потребуют другого набора знаний, умений, навыков и компетенций. Чтобы хоть как-то компенсировать этот недостаток, в новом стандарте Министерства образования и науки имеется фраза: «Выпускники могут осуществлять профессиональную деятельность в других областях профессиональной деятельности и сферах профессиональной деятельности при условии соответствия уровня их образования и полученных компетенций требованиям к квалификации работника». Очевидно, что такая формулировка стандарта вызывает массу вопросов при ее практическом использовании.

Перечисление в стандартах образования областей профессиональной деятельности и соответствующих им профессиональных стандартов, с одной стороны, существенно ограничивает потенциальные возможности трудоустройства выпускников, и, с другой стороны, оставляет без специалистов с необходимой квалификацией большое количество уже существующих или потенциальных рабочих мест, для которых профессиональные стандарты либо еще не написаны, либо вообще никогда написаны не будут. Если вузы сосредоточатся только на профессиональной подготовке в соответствии с требованиями вводимых стандартов, то это приведет либо к необходимости создания дополнительной системы послевузовской переподготовки, либо к замещению вакантных должностей сотрудниками с недостаточными с точки зрения профессиональных требований компетенциями.

Применительно к предмету обучения – Прикладной информатике – отметим, что, с точки зрения интересов общества и государства, целью подготовки являются интересы всей информационной отрасли, а не двух выделенных новым стандартом областей. Известно [2], что информационная отрасль включает в себя два сектора: первичный и вторичный (рис. 1). В первичном секторе информационная деятельность является главенствующей формой деятельности, а ее основным продуктом являются информация и знания, информационные средства, системы и технологии, информационные услуги.

Вторичный сектор информационной отрасли входит в состав других структур экономики, связанных с жизнедеятельностью человека, материальным производством и производством энергии. Информационная деятельность и информационные технологии в этом секторе направлены на повышение эффективности и производительности отраслей второго информационного сектора, то есть информационные товары и услуги, производимые в первичном и вторичном секторах, потребляются внутри вторичного сектора.

Традиционно выпускники направления Прикладная информатика находили себе рабочие места как в первом, так и во втором информационных секторах. Их наличие определяло специ-

фику обучения. Рассматривая всю информационную отрасль как единую научную область знаний и обеспечивая подготовку с учетом ее структуры, мы можем подготовить специалиста, который востребован в самых разных ее сферах. Очевидно, что, если вуз ведет обучение обеспечивая фундаментальность образования своего выпускника, он дает ему полное представление о текущем состоянии соответствующей области знаний. В этом случае, что проверено многолетним опытом, выпускнику будет гораздо проще адаптироваться к специфическим задачам его конкретного рабочего места на любой позиции информационной отрасли.



Рис. 1. Структура информационной отрасли экономики по Р. М. Юсупову

В случае профессиональной подготовки требования к содержанию образования в области Прикладной информатики в новом стандарте оказываются существенно более низкими по сравнению с существующими. Ориентируясь на них разработчик учебного плана и содержания обучения должен бороться с искушением максимально удешевить и упростить подготовку за счет исключения непрофильных для конкретной профессиональной области разделов. Фактически речь идет о том, что требования стандарта образования не принуждают вуз давать своему выпускнику полное представление об интересующей его области знаний (в данном случае информатике) [3].

Таким образом, если новый стандарт будет реализован, мы имеем шанс потерять системность и фундаментальность обучения. В результате новые стандарты образования будут существенно ограничивать набор возможных и необходимых экономике областей профессиональной деятельности, а также возможности выпускников как в части удовлетворения своей потребности в образовании, так и в части дальнейшего трудоустройства. Учитывая предстоящие в этом случае затраты на переподготовку и дальнейшее повышение квалификации выпускника в результате реализации новых стандартов образования мы имеем шанс оказаться в существенном проигрыше.

Заключение

Внедрение в практику высшей школы новых стандартов высшего образования, основанных на профессиональных стандартах министерства труда и социальной защиты, создает угрозу потери фундаментальности результатов обучения. Если вузы сосредоточатся на профессио-

нальной подготовке, например, в области только связи, информационных и коммуникационных технологий, то необходимыми экономике специалистами будут обеспечена только относительно небольшая часть экономики. Как следствие, позднее в стране возникнут проблемы с подготовкой кадров высшей квалификации, потребуются дополнительные затраты на создание системы переподготовки, что, в конечном счете, приведет к снижению общего уровня образования в стране.

ЛИТЕРАТУРА

1. Семенова Е. Г. Оценочные средства как показатели качества компетенций выпускников высшей школы / Е.Г. Семенова, А.Г. Степанов // Актуальные проблемы экономики и управления. 2015. №3 (7). С. 53 – 66.
2. Юсупов Р.М. Информационные технологии и экономика информационного общества. Инновации. / Р. М. Юсупов // 2013. № 11 (181). С. 40 – 46.
3. Заболотский В.П. Информатика как предмет обучения в высшей школе / В.П. Заболотский, А.Г. Степанов, Р.М. Юсупов // Труды СПИИРАН. 2004. Т. 1. № 2. С. 316 – 347.

A. G. Stepanov (St. Petersburg State University of Aerospace Instrumentation (GUAP))

Professional preparation or education?

Abstract.The problem of the possible loss of the fundamentality of education in the implementation of new federal state standards of higher education based on the professional standards of the Ministry of Labor and Social Protection

ВОПРОСЫ ОБЕСПЕЧЕНИЯ КАЧЕСТВА ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА В СОВРЕМЕННЫХ УСЛОВИЯХ

А. М. Боронахин[1], **А. Н. Ткаченко**[2], **Р. В. Шалымов**[3]
(*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина), 197376, С.-Петербург, ул. Профессора Попова, 5, (812) 234-30-59, fibs@etu.ru*)

Аннотация

Образовательный процесс является сложным и многофакторным и затрагивает большое количество лиц и организаций. Поэтому обеспечение его качества – это задача, решение которой в современных условиях невозможно без использования системного подхода. Первым шагом в этом направлении всегда становится определение с самим пониманием качества, выработка критериев качественного образования. После необходимо сформировать методику и осуществлять регулярный контроль качества по этим критериям. На основе результатов такого мониторинга должен проводиться анализ наиболее значимых внутренних и внешних факторов, влияющих на процесс. Последним шагом здесь становится прогноз развития и выработка рекомендаций для органов управления образовательной организацией.

Одной из сложностей осуществления системного подхода к обеспечению качества образования является необходимость консолидации усилий разных сторон этого процесса. Однако стоит учитывать, что все они имеют свои интересы, не во всех аспектах коррелирующие между собой.

В центре внимания настоящей работы находятся анализ проблемы организации взаимодействия между участниками образовательного процесса, с учетом их интересов и имеющегося опыта; преимущества и недостатки существующих подходов к оценке качества образования; использование методов управления качеством для совершенствования инженерного образования в условиях динамично меняющихся требований в современных условиях.

Ключевые слова: качество образования, компетенция, образовательный процесс, управление качеством.

Введение

Вопрос о соответствии качества образования в ВУЗе требованиям заказчика является сложным и неоднозначным: в настоящее время различны как методики оценки, так и подходы к управлению качеством образования, применяемые каждой отдельной образовательной организацией. Даже само понятие «заказчик» несет размытую нагрузку: это студенты и их родители, работодатели, Министерство образования и науки, общество в целом.

Качество образования в значительной мере определяется степенью удовлетворения ожиданий и запросов всего многообразия потребителей: для студента это будут мотивация, интерес к учебному процессу, востребованность на рынке труда и вероятность успешного трудоустройства, для родителя – стабильное будущее ребенка, для работодателя – приобретение квалифицированного сотрудника, для министерства образования – выполнение образовательных стандартов, для общества – формирование адаптированной к социуму личности, способной применять полученные при обучении навыки и умения для повышения уровня и качества жизни [1]. Таким образом, образовательная организация, стремящаяся соответствовать всем требованиям, вынуждена выстраивать системное взаимодействие со всеми сторонами, участвующими в образовательном процессе (в том числе, находящимися на входе и выходе этой системы). Простое рассмотрение пары школа-ВУЗ или ВУЗ-работодатель не в состоянии решить все возникающие задачи, т.к. у каждой из перечисленных сторон имеются свои интересы, которые не всегда в полной мере совпадают [2].

Современные реалии налагают на образовательный процесс еще одно условие, особенно ярко проявляющееся при подготовке студентов технических направлений: высокая скорость развития в ряде технических и технологических отраслей, особенно в области информационных технологий, что предполагает столь же быстрое перестроение содержания читаемых дисци-

[1] Доктор технических наук, декан факультета информационно-измерительных и биотехнических систем (ФИБС), профессор кафедры лазерных измерительных и навигационных систем (ЛИНС).
[2] Кандидат технических наук, доцент кафедры лазерных измерительных и навигационных систем (ЛИНС).
[3] Кандидат технических наук, доцент кафедры лазерных измерительных и навигационных систем (ЛИНС).

плин, а порой и всей структуры образовательных программ. Это приводит к необходимости повышения квалификации преподавателей в наиболее передовых областях современной науки и техники, включая не только их прямую специальность, но и смежные направления подготовки (изучение современных информационных пакетов, циклов производства и эксплуатации технических систем, включая их дальнейшую утилизацию и т.д.).

В работе рассматриваются вопросы обеспечения качества инженерного образования в условиях динамически меняющихся требований. Приводятся классификация и анализ влияющих на образовательный процесс факторов, процедуры оценки качества подготовки выпускников. В роли элемента системы внешней оценки выступают профессиональные стандарты. Затрагивается проблема интеграции профессиональных стандартов и федеральных государственных образовательных стандартов.

Системное взаимодействие: интересы разных сторон образовательного процесса

В соответствии с Федеральным законом об образовании в Российской Федерации «Образование – единый целенаправленный процесс воспитания и обучения, являющийся общественно значимым благом и осуществляемый в интересах человека, семьи, общества и государства, а также совокупность приобретаемых знаний, умений, навыков, ценностных установок, опыта деятельности и компетенции определенных объема и сложности в целях интеллектуального, духовно-нравственного, творческого, физического и (или) профессионального развития человека, удовлетворения его образовательных потребностей и интересов». [3] Исходя из такой формулировки, можно сделать несколько заключений:

- процесс образования осуществляется в интересах целой группы потребителей («человека, семьи, общества и государства»);
- понятие образования крайне широко и охватывает воспитание, обучение, приобретение компетенций, удовлетворение потребностей и многое другое;
- образование следует понимать, как «единый целенаправленный процесс», т.е. правильное построение процесса требует единения усилий со стороны всех участников для достижения поставленной цели.

Следует, однако, отметить, что, учитывая многогранность понятия «образование» и значительное количество заинтересованных сторон, создание в полной мере «единого целенаправленного процесса» на практике является весьма затруднительным по причине различного взгляда на итоговый результат со стороны каждого из участников этого.

Участниками образовательной системы являются такие организации, как школы, ВУЗы и училища, предприятия, государство. Интересы и место в системе у них разные.

Интересом со стороны государства является адаптация личности к социуму, наполнение приоритетных направлений развития страны квалифицированными кадрами. Блюдя свои интересы, государство выступает в роли контролера деятельности остальных субъектов образовательной системы, а также влияет как на количество выпускников, получающих ту или иную специальность (в первую очередь через распределение контрольных цифр для бюджетного приема), так и на содержание образовательных программ (через федеральные образовательные стандарты (ФГОСы)).

Интерес предприятий – получение наиболее квалифицированных кадров, при этом для них важным является наличие у трудоустраивающегося тех компетенций, которые востребованы именно на предлагаемой ему вакансии. Таким образом, хотя каждое предприятие хочет иметь квалифицированных сотрудников, однако, взгляд на набор компетенций, эту квалификацию формирующих у них может быть различный (тем более различный, чем более далеки друг от друга области их деятельности). В образовательной системе предприятия выступают в первую очередь как выходное звено, чего зачастую недостаточно для достижения интересов предприятия: приходится проводить предварительный отбор, а затем доучивать поступающие на вход предприятия кадры, т.к. найти «идеального» будущего сотрудника практически невозможно. Ряд предприятий включаются в образовательный процесс высших и средних образовательных учреждений. Степень этого включения может быть различной, но призвана реализовать стремление предприятия получить на входе в свою структуру максимально подготовленного выпускника.

Интерес со стороны организаций высшего или среднего профессионального образования состоит в их востребованности их абитуриентами на фоне конкурентов, а также в высокой оценке со стороны государственных контролирующих органов и предприятий (прохождение аккредитации, получение признания и статуса). Факторами, выделяющими ВУЗ или училище на фоне конкурентов, являются как качество поступающих в него абитуриентов (средний балл ЕГЭ, количество победителей олимпиад и конкурсов различного уровня), так и статус его недавних выпускников (уровень их зарплат, карьерный рост, востребованность на рынке труда). Помимо обязательной деятельности, подразумевающей обеспечение процесса обучения, установленного соответствующим ФГОСом, сильнейшие учебные организации этого уровня для достижения превосходства над конкурентами дополнительно используют такие инструменты, как ранняя профориентационная деятельность в среде школьников (проведение олимпиад, конкурсов, ведение кружков, привлечение школьников к проектной деятельности), укрепление взаимоотношений с предприятиями сектора (проведение совместных НИР и ОКР, повышение квалификации, целевой прием).

Интерес со стороны школы – это также ее место в рейтинге школ района, города, области, страны. Оценивают школу, как государство, так и родители, когда отдают туда своих детей на обучение. Среди многих других факторов одним из важнейших, особенно в старшей школе, является статус ее выпускников (средний балл ЕГЭ, количество победителей олимпиад и конкурсов различного уровня, престижность ВУЗов, в которых они учатся). Лучшие школы проводят собственную или совместную с ВУЗами-партнерами профориентационную деятельность, нацеленную на ознакомление с будущей специальностью и поднятие мотивации школьников для достижения успехов в учебе. [2]

Исходя из опыта ведущих представителей каждой из перечисленных групп можно сделать следующее заключение: многие действия, осуществляемые ими, находятся в одном векторе, но обычно проводятся независимо друг от друга. На какой же основе возможно объединить усилия всех перечисленных организаций?

Современная политическая и экономическая ситуация в России и мире формирует новые вызовы перед промышленностью в нашей стране. Важнейшей целью развития отраслей промышленности Российской Федерации становится формирование и сохранение глобальной конкурентоспособности. Это должно с одной стороны обеспечить импортозамещение в жизненно важных для сохранения государственного суверенитета областях, а с другой – выход отечественных товаров на мировой рынок, привлечение капитала в страну. Стоит, однако, отметить, что развитие технологического превосходства потребует создания центров притяжения с участием ведущих ученых не только России, но и мира. Привлечение людей, создающих современные тренды в науке и технике, позволит направить усилия промышленности и образования по наиболее перспективному пути.

Оценка мировых тенденций, получаемая при анализе направлений работ ведущих международных научных коллективов, позволит реорганизовывать образовательную систему в соответствии с текущей ситуацией. Перспективные технические направления будут оформлены в англоязычные образовательные программы при ведущих ВУЗах для дальнейшего привлечения к ним лучших мировых ученых. На основе подобных образовательных программ потребуется создать систему повышения квалификации сотрудников ВУЗа и представителей предприятий отрасли. Таким образом, вокруг перспективного направления развития науки и техники будет сформирован образовательный кластер. Примерный состав такого кластера будет иметь вид: несколько предприятий, ведущих работу в этой области; один-два ВУЗа обеспечивающих подготовку специалистов высокой квалификации для обозначенной отрасли; пару образовательных организаций среднего профессионального образования, позволяющих обеспечить предприятия требуемыми кадрами рабочих квалификаций; несколько школ. Формирование системного взаимодействия между всеми сторонами процесса потребует наладить специализированную подготовку в учебных заведениях каждого уровня (с учетом всех последующих этапов, вплоть до трудоустройства). В образовании, построенном таким образом, неизбежно должно осуществляться глубокое взаимодействие между сторонами процесса:

- участие преподавателей ВУЗа, училища и представителей промышленности в школьном учебном процессе (ранняя профориентация, адаптация учебных планов, техническое

оснащение, привлечение специалистов для ведения занятий, проведение специализированных «кружков»); [4]

- обеспечение для выпускников училища возможности повышения квалификации за счет продолжения обучения в ВУЗе;
- участие предприятия в учебном процессе ВУЗа и училища (профориентация студента, использование целевого приема, согласование учебных планов (вплоть до специализированных образовательных программ), создание базовых кафедр, техническое оснащение, привлечение специалистов предприятия для проведения занятий);
- система повышения квалификации (как для представителей промышленности и школьных учителей в ВУЗе, так и для преподавателей/учителей на предприятии).

Важным фактором здесь является подготовка учащегося на текущей ступени образования к переходу на следующую. При этом необходима не только организация дополнительной специальной подготовки, позволяющей в дальнейшем легче и быстрее адаптироваться в новой среде, но и постоянно присутствующая профориентационная и мотивационная составляющая в процессе образования. Важно, чтобы учащийся четко видел имеющиеся возможности и понимал для чего делается тот или иной шаг в его обучении (где ему в дальнейшем пригодятся полученные знания и навыки). [5]

Создание полноценного образовательного кластера по такой модели является сложной задачей. Как правило, за один шаг решить ее не представляется возможным. В этом случае стоит двигаться поступательно, укрепляя взаимодействие между отдельными сторонами процесса. Примерами таких шагов, пройденных СПбГЭТУ «ЛЭТИ», могут стать организация попечительского совета факультета информационно-измерительных и биотехнических систем, позволяющая укрепить взаимодействие ВУЗа и предприятий в конкретной области [5]; организация школы-технопарка МОБУ «СОШ «ЦО «Кудрово», создающая базу специализированно подготовленных абитуриентов для ВУЗа [6]; организация учебно-методологического центра «Утилизация и обезвреживание отходов», направленная на консолидацию работы науки, образования и промышленности в этом направлении [5]; осуществление индивидуальной или групповой подготовки по принципам CDIO [7] и многое другое.

Качество образовательного процесса: подходы к оценке

В соответствии с уже упомянутым Федеральным законом об образовании в Российской Федерации «Качество образования – это комплексная характеристика образовательной деятельности и подготовки обучающегося, выражающая степень их соответствия федеральным государственным образовательным стандартам, образовательным стандартам, федеральным государственным требованиям и (или) потребностям физического или юридического лица, в интересах которого осуществляется образовательная деятельность, в том числе степень достижения планируемых результатов образовательной программы». Уже здесь видно, что качество образования это, с одной стороны, – степень соответствия государственным требованиям, а с другой, – степень удовлетворения потребностей конкретных лиц и организаций.

Так что же такое качество образования и как его оценить? Это вопрос, на который разные категории участников образовательного процесса ожидаемо дадут различные ответы: все ждут выполнения своих ожиданий и запросов.

Среди критериев качества высшего образования называют такие как структура и качественный состав профессорско-преподавательского состава, качество учебно-методического обеспечения, материально-техническую базу ВУЗа, интеллектуальный потенциал учебного заведения (количество НИР, число защит, публикационная активность и др.), численность и качественный состав студентов (балл ЕГЭ на входе, текущая и итоговая успеваемость), а также уровень выпускников (численность, уровень трудоустройства и карьеры). [8] Однако потребителю образовательного процесса многие из рассматриваемых критериев могут показаться незначительными (вряд ли поступающий в ВУЗ будет оценивать качество образования, которое он может там получить, по количеству докторов наук, защитившихся в текущем году).

Если взглянуть на образовательный процесс глазами обучающегося, то качество получаемого/полученного образования будет восприниматься им с различных позиций на каждом из эта-

пов жизни. В течение обучения помимо конечного результата его будет волновать комфортность, обеспеченность, интересность процесса, ощущение мотивации к саморазвитию, творческая атмосфера и многое другое. [2, 5] Сразу по окончании обучения он будет оценивать качество образования по своей востребованности на рынке труда и успешности своего трудоустройства (уровень зарплаты, интересная работа, возможности к карьерному росту), соответствию полученных им знаний текущей ситуации в отрасли. С течением времени качество полученного в прошлом образования оценивается как по пригодившимся на работе навыкам и умениям, полученным на студенческой скамье, так и по уровню развития способности к самообразованию (как для движения по карьерной лестнице, так и для возможности сменить рабочее место). Оценки качества образования в этом случае всегда будут субъективны и могут анализироваться и учитываться только при обработке мнений статистически значимого количества респондентов.

Потребители продукта образовательного процесса (предприятия, государство и общество) оценивают его с других позиций и, как правило, более формализованных. Предприятие ищет себе сотрудников, удовлетворяющих целому ряду конкретных характеристик, которые могут быть оформлены в профессиональные стандарты. Тогда оценкой качества образования становится соответствие выпускников этим стандартам. С другой стороны, государство, в лице Минобрнауки, оформляет свое видение качества в виде официальных инструкций и распоряжений (ФГОСы), являющихся обязательными к исполнению. Несоответствие государственным образовательным стандартам может приводить к неаккредитации направлений подготовки. Этот подход позволяет контролировать только качество обучения (а не всего процесса образования) и не содержит большого числа градаций оценки этого качества (фактически позволяет лишь отсеивать самые слабые учебные заведения или их отдельные направления подготовки). При этом следует учитывать, что до настоящего момента требования государства не имели систематической связи с требованиями промышленности. Такое несоответствие могло приводить к сложностям при попытках построения ВУЗами совместных образовательных программ или создании специализированных образовательных траекторий в рамках уже существующих программ. Ожидаемым шагом в направлении сближения ФГОСов и профессиональных стандартов должен стать ФГОС 3++, предусматривающий в числе прочего определение области профессиональной деятельности в соответствии с реестром профессиональных стандартов, изменение подхода к дисциплинам (модулям) по выбору, формирование перечня профессиональных компетенций для программ бакалавриата на основе профессиональных стандартов.

Управление качеством образования: инженерное образование в условиях динамически меняющихся требований

Закон об образовании в Российской Федерации относит к компетенции, правам, обязанностям и ответственности образовательной организации в числе прочего «проведение самообследования, обеспечение функционирования внутренней системы оценки качества образования». Система такого внутреннего мониторинга, направленная на обеспечение соответствующей информацией управляющих органов образовательной организации, является важным условием прохождения государственной аккредитации.

Каждая образовательная организация самостоятельно определяет методику внутренней оценки качества, поэтому как используемые критерии, так и процедуры мониторинга могут значительно отличаться, несмотря на то, что преследуемые цели, как правило, совпадают:

- своевременное определение наиболее значимых факторов, влияющих на качество образования в настоящее время;
- получение информации о тенденциях к изменению в системе образования и причинах этих изменений, прогнозирование развития образовательной системы;
- принятие обоснованных управленческих решений по видоизменению отдельных аспектов или всего подхода к процессу образования в организации.

Количество методов управления качеством и их вариаций велико [9,10]. При этом многие из них схожи и поддаются группировке. Можно выделить, например, такие три подхода (группы методов), как [1]:

1. оценочный метод управления качеством (систематическое проведение самооценки, направленное на выявление внутренних и внешних влияющих факторов, разделяемых на четыре категории: сильные стороны, слабые стороны, возможности, угрозы);
2. метод, основанный на принципах всеобщего управления качеством (нацелен на непрерывное повышение степени реализации требований клиентов, значения финансовых показателей организации и уровня удовлетворенности служащих компании своей работой);
3. метод, основанный на требованиях международных стандартов ИСО 9001:2015 (предполагает определение заинтересованных сторон, выявление их требований к качеству услуг, применение процессного подхода для совершенствования деятельности организации в сторону достижения выявленных требований) [11].

Результатом работы системы управления качеством образования вне зависимости от конкретного подхода, лежащего в ее основе является изменение структуры и содержания образовательного процесса таким образом, чтобы его качество улучшалось (или хотя бы не снижалось). При этом следует учитывать, что в зависимости от принятого подхода, выбранных критериев, а также используемых инструментов изменения текущей ситуации, из одного и того же состояния можно переместиться в одно из соседних состояний, характеризующихся лучшим качеством, большим количеством способов.

Одним из главных вызовов современности, встающих перед инженерным образованием является скорость научно-технического прогресса. Поддержание высокого качества образования в таких условиях предполагает постоянный контроль изменений не только в тех отраслях, по которым учебное заведение осуществляет выпуск специалистов, но и в смежных областях. В противном случае выпускник будет не востребован на рынке, а значит и эффективность образования встанет под вопрос. Рассмотрим эту ситуацию на примере студента, обучающегося по направлению «Приборостроение» [12]. Помимо классических для этого направления подготовки знаний и умений в области электротехники, метрологии, физических основ получения информации, проектирования и технологии производства приборов и систем и методов обработки измерительной информации выпускнику полезны, а порой и необходимы сопутствующие компетенции. Среди таковых в современных реалиях стоит выделить:

- знакомство с основными типами используемых на производствах и в научных исследованиях специализированных информационных пакетов, позволяющих осуществлять макетирование, моделирование, расчеты и др.;
- понимание всего цикла разработки и производства технической системы (от идеи, через разработку макета, до серийного выпуска), включая конечный этап утилизации изделия;
- знание основ систем управления, с которыми измерительная техника часто работает в связке;
- осведомленность о достижениях в области материаловедения и многое другое.

Помимо приведенных «технических» компетенций выпускнику могут пригодиться и знания, полученные в области маркетинга, юриспруденции и экономики предприятия. Фактически на выходе из учебного заведения выпускник этого направления должен иметь возможность как работать на уже существующем предприятии отрасли в любом из отделов (конструктором, технологом, метрологом и т.д.), так и организовать собственное дело.

При этом следует отметить, что преподаватели, осуществляющие подготовку также должны постоянно поднимать свою квалификацию не только в области читаемых ими дисциплин и проводимых научных исследований (что само собой разумеется), но и в смежных отраслях науки и техники. Поэтому повышение качества образования в техническом ВУЗе в современных реалиях невозможно без полноценной системы повышения квалификации преподавателей. Ее поддержание по всем знаковым направлениям обычно не представляется возможным только лишь за счет привлечения своих специалистов (качественная образовательная система не может быть замкнутой). Зачастую необходимым является не только привлечение экспертов из других ВУЗов, но и специалистов с предприятий.

Поэтому система управления качеством должна позволять в числе прочих принимать такие управленческие решения, которые позволят с одной стороны отслеживать все современные тенденции в интересующих отраслях и вокруг них (на основании чего могут приниматься решения не только о модернизации существующих направлений подготовки, но и об их закрытии

или открытии новых), а с другой – вносить соответствующие текущей ситуации изменения в образовательные программы с одновременным повышением квалификации преподавателей, читающих соответствующие дисциплины. Такой подход может быть реализован, например, при наличии в ВУЗе возможности для студента двигаться по различным образовательным траекториям (вплоть до индивидуальных). [6,4] Тогда небольшой группе студентов, имеющих соответствующие наклонности можно давать возможность углублять свое образование, например, в области информационных технологий, а другой – в области утилизации отходов.

Следует также отметить, что, несмотря на стремление к идеальному случаю «опережающего образования» когда знания, получаемые на студенческо скамье, являются перспективными и лишь в ближайшем будущем (к моменту выпуска) станут действительно востребованными на рынке, предусмотреть развитие науки и техники на весь период работоспособности выпускающегося специалиста является невозможным. Поэтому на фоне прочего не стоит забывать, что одним из важнейших факторов качественного образования овладение навыком самообразования. Возможность в дальнейшем самостоятельно повышать свою квалификацию для инженерного специалиста предполагает с одной стороны крепкие базовые знания, на которые, как на фундамент могут быть наложены новые, а с другой стороны получение достаточно разнородного опыта в стенах ВУЗа, чтобы новые компетенции могли быть увязаны с какими-либо из уже полученных.

Опыт факультета информационно-измерительных и биотехнических систем «ЛЭТИ»: система менеджмента качества в бакалавриате

Образование адекватное современным реалиям должно одновременно обеспечивать очень разноплановые возможности. С одной стороны, необходимым в условиях быстроменяющихся требований является получение сильной «базовой» подготовки и навыков самообразования. С другой – желание стать востребованным специалистом сразу после выпуска можно обеспечить только углубленной специализацией под конкретную группу работодателей. Эти на первый взгляд разнонаправленные устремления можно объединить с использованием системных взаимодействий в образовании и обратных связей по качеству получаемых компетенций.

Именно с этой целью на факультете информационно-измерительных и биотехнических систем (ФИБС) СПбГЭТУ «ЛЭТИ» уже несколько лет ведется планомерная работа по формированию полноценной системы контроля качества образования по своим направлениям подготовки (Биотехнические системы и технологии, Приборостроение, Техносферная безопасность). Схема организации процесса представлена на рисунке 1.

Базовая подготовка будущих специалистов начинается еще в школе, где благодаря различным проектам взаимодействия (например, влиянию на учебные планы в центре образования «Кудрово») становится возможным усилить ту или иную составляющую обучения (ввести факультативные занятия по физике, организовать проектную работу школьников и многое другое). Определение текущих компетенций школьных выпускников осуществляется входным тестированием по ключевым фундаментальным дисциплинам, читаемым институтом фундаментального инженерного образования (ИФИО) СПбГЭТУ «ЛЭТИ» на первых двух курсах ФИБС. По результатам этого тестирования, помимо задействования обратной связи, позволяющей изменить качество подготовки абитуриента, ИФИО осуществляет подготовку младшекурсников в соответствии с уровнем их текущих знаний базовых дисциплин (корректируется содержание отдельных разделов базовых дисциплин, организуются бесплатные дополнительные занятия).

По окончании второго курса наступает пора изучения специальных дисциплин, касающихся конкретного направления подготовки. По окончании третьего курса студенты проходят через два ключевых события, позволяющих скорректировать их образовательные траектории. Первым является тестирование остаточных знаний по ряду базовых и специальных дисциплин, имеющее целью определить текущий уровень квалификации и предрасположенность к определенным видам технической деятельности. Вторым становится факультетское соревнование студенческих проектов «TechnoStart». Задания для студентов заранее формируются представителями предприятий-партнеров, входящих в попечительский совет факультета. Работа в командах над совместным проектом позволяет каждому из студентов почувствовать свои сильные и

слабые стороны в разных видах технического творчества. По окончании такого двухэтапного процесса самоопределения выявляются студенты, имеющие определенные таланты: конструктора, технолога, инженера, метролога, исследователя. Для каждой из перечисленных групп на четвертом курсе читаются свои дополнительные дисциплины, организованные в рамках вариативной части. Попечительский совет, отслеживая подготовку ребят на четвертом курсе, участвует как в формировании актуальных тематик для будущих ВКР (выдаются уже на третьем курсе), так и в оснащении лабораторий необходимым, для получения требуемых компетенций оборудованием. По окончании процесса обучения и защиты студентами ВКР представители предприятий, работающие в составе государственных аттестационных комиссий (ГАК) факультета, собираются на общее заседание и оценивают качество образования у защитившихся, указывают на имеющиеся недостатки и пути совершенствования. В соответствии с их рекомендациями могут быть откорректированы рабочие программы для последующих поколений студентов.

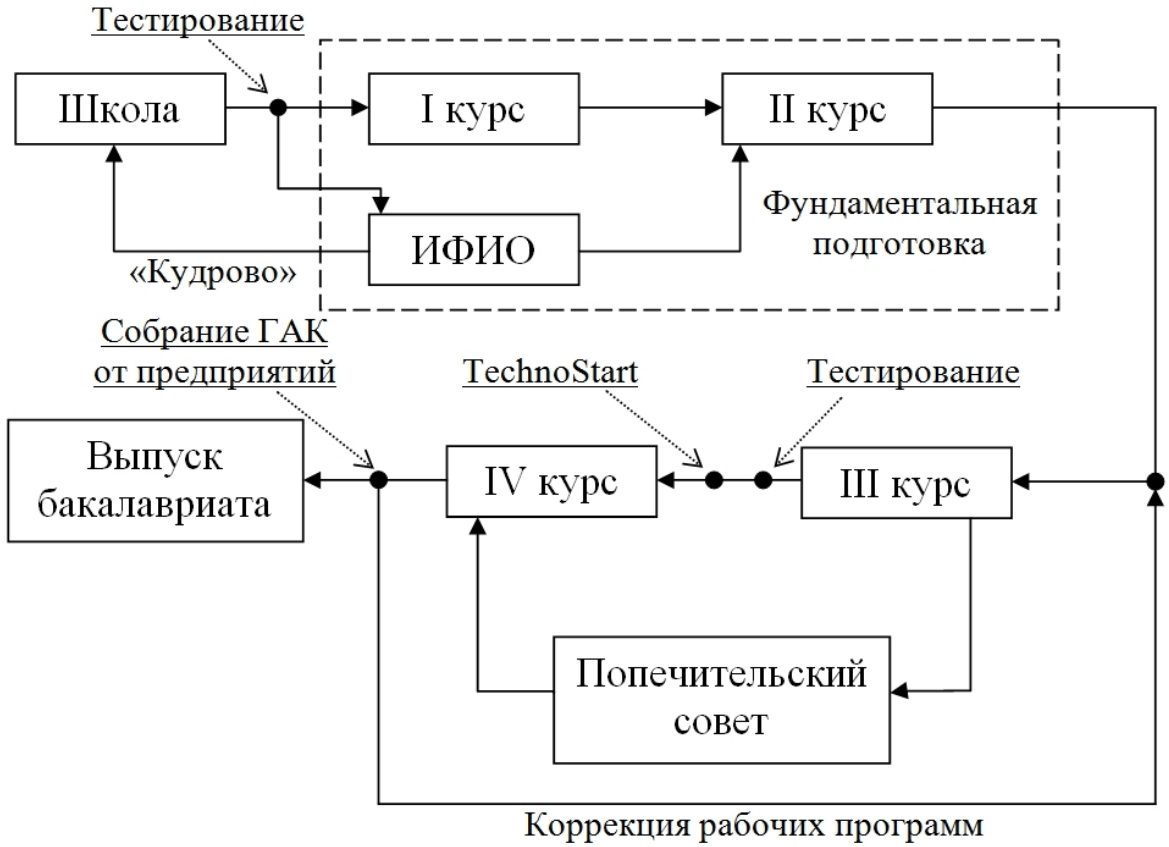


Рис. 1. Организация образовательного процесса на ФИБС СПбГЭТУ «ЛЭТИ»

Заключение

Обеспечение образования высокого качества является задачей, которую стремятся решить все стороны образовательного процесса. Тем не менее, в результате как различного понимания самого термина «качество образования», так и приверженности к разным методам его совершенствования, разные участники процесса зачастую действуют не согласованно. Однако очевидным является, что без системного подхода с формированием образовательных кластеров или подобных структур достичь требуемых значений по многим критериям качественного образования становится практически невозможно.

Сложной задачей является также определение значимости влияющих факторов и формирование системы критериев качества образования. Обязательным к исполнению, в должной мере проработанным и относительно легко контролируемым является подход к обеспечению качества образования через государственный контроль (лицензирование и аккредитацию). Однако этот наиболее широко используемый инструмент имеет существенный недостаток: он позволяет оценить качество фактически по двухбалльной системе (удовлетворительно или не удовле-

творительно), в то время, как стоящая в настоящее время перед каждой организацией и всей образовательной системой в целом задача может быть сформулирована, как «максимизация качества образования в текущих внешних условиях».

В таких условиях жизненно необходимым для каждой образовательной организации является формирование системы мониторинга качества образования. Эта система должна позволять анализировать результаты образовательной деятельности, выявлять наиболее значимые ее аспекты и тенденции к их изменениям, а также разрабатывать прогноз, на основании которого должны приниматься управленческие решения в организации.

Повышение качества образования требует значительных усилий от всех заинтересованных сторон.

Работа выполнена при поддержке гранта Президента Российской Федерации по государственной поддержке ведущих научных школ Российской Федерации НШ-4165.2018.8.

ЛИТЕРАТУРА

1. Мазитов М.А. Система менеджмента качества в образовании / М.А. Мазитов, Р.Ф. Сагитов. // ВЕСТНИК ОГУ. - 2011. - № 2 (121). - С. 205-208.
2. Боронахин А. М. Построение неразрывного образовательного процесса на факультете информационно-измерительных и биотехнических систем СПбГЭТУ «ЛЭТИ» с целью повышения мотивации всех участвующих сторон / А. М. Боронахин, Р. В. Шалымов // Современное образование: содержание, технологии, качество. - 2017. - Т. 2. - С. 176-178.
3. Федеральный Закон от 29 декабря 2012 г. № 273-ФЗ «Об образовании в Российской Федерации».
4. Боронахин А.М. Междисциплинарность при практико-ориентированной подготовке бакалавров в соответствии с подходом CDIO / А.М. Боронахин, А.А. Минина, Р.В. Шалымов // Инженерное образование, № 20, 2016 – М.: Ассоциация инженерного образования России, 2016. – С. 159-163.
5. Боронахин А.М. Использование возможностей стратегического партнерства с предприятиями при построении образовательного процесса на факультете информационно-измерительных и биотехнических систем СПБГЭТУ "ЛЭТИ" / А.М. Боронахин, Д.Ю. Ларионов, Л.Н. Подгорная, Р.В. Шалымов // Планирование и обеспечение подготовки кадров для промышленно-экономического комплекса региона. 2017. Т. 1. С. 31-34.
6. Боронахин А.М. Особенности реализации учебного процесса в рамках усиления инженерной практико-ориентированной подготовки / Боронахин А.М., Минина А.А., Подгорная Л.Н., Шалымов Р.В. // Материалы XV Всероссийской научно-практической конференции «Планирование и обеспечение подготовки кадров для промышленно-экономического комплекса региона», Т. 1, 2016 – С.-Петербург: СПбГЭТУ «ЛЭТИ», 2016. – С. 65-70.
7. Боронахин А.М. Организация практико-ориентированной лаборатории информационно-измерительных и биотехнических систем / А.М. Боронахин, А.А. Минина, Р.В. Шалымов // Материалы XIV Всероссийской научно-практической конференции «Планирование и обеспечение подготовки кадров для промышленно-экономического комплекса региона», №1, 2015 – С.-Петербург: СПбГЭТУ «ЛЭТИ», 2015. – С. 166-171.
8. Егорошин А.П. Менеджмент, маркетинг и экономика образования: учеб. Пособие, 2-е изд., перераб. / Под ред. А.П. Егорошина, Н.Д. Никанорова. Н. Новгород, 2004.
9. Варченко Е. И. Методы управления качеством образования в современной школе // Молодой ученый. - 2013. - №4. - С. 541-544.
10. Прохоров Ю.К. Управление качеством: Учебное пособие. – СПб: СПбГУИТМО, 2007. – 144 с.
11. ГОСТ Р ИСО 9001-2015 от 01 ноября 2015 г. «Системы менеджмента качества. Требования».
12. А.М. Боронахин, Р.В. Шалымов СПбГЭТУ «ЛЭТИ»: Приборостроение // Техносфера№2, 2016. С.25-26.

A.M. Boronachin, A.N. Tkachenko, R.V. Shalymov (Saint Petersburg Electrotechnical University "LETI", Saint Petersburg)

Вопросы обеспечения качества образовательного процесса в современных условиях

The educational process is complex and multifactorial and involves a large number of individuals and organizations. Therefore, ensuring its quality is a task, the solution of which in modern conditions is impossible without the use of a systematic approach. The first step in this direction is always the establishment of the very concept of quality, the development of criteria for quality education. After that, it is necessary to formulate a methodology and carry out regular quality control on these criteria.

Based on the results of such monitoring, the most significant internal and external factors affecting the process should be analyzed. The last step here is to forecast development and develop recommendations for the management of the educational organization.

One of the complexities of implementing a systematic approach to ensuring the quality of education is the need to combine the efforts of various parties in this process. After all, they all have their own interests, which do not completely correlate with each other.

The focus of this work is an analysis of the problem of organizing interaction between participants in the educational process, taking into account their interests and experience; advantages and disadvantages of existing approaches to assessing the quality of education; use of quality management methods to improve engineering education in the face of rapidly changing requirements.

ПРИМЕНЕНИЕ ПРОЕКТНОЙ ДЕЯТЕЛЬНОСТИ ПРИ ОБУЧЕНИИ В ВУЗЕ

Р. Р. Фаткиева
(*Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский институт информатики и автоматизации» Российской академии наук,
199178, С.-Петербург, 14 линия, дом 39, +7-(812)-328-43-69, rikki2@yandex.ru*)

К. С. Константинова
*(Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина),
197376, С.-Петербург, ул. Профессора Попова, дом 5,+7-(812)-234-46-51,lisoffski@yandex.ru)*

Аннотация

Проведен анализ этапов проектной деятельности в ВУЗе. Рассмотрены вопросы оценивания проектной деятельности на каждом из этапов. Предложен метод оценки проектной деятельности, позволяющий произвести оценку совместной и индивидуальной деятельности студентов. Предложенный метод позволяет осуществить прогноз развития проекта и осуществить планирование мероприятий для его совершенствования.

Ключевые слова: учебная деятельность, научно-исследовательская работа учащихся, совместная деятельность, проектная деятельность.

Введение

Происходящие в современном обществе изменения требуют использование новых подходов к обучению, нацеленных на индивидуальное развитие личности, творческую инициативу, выработку навыка самостоятельной работы в условиях увеличивающегося объема информации с одной стороны и необходимостью достижения синергетического эффекта от работы в команде с другой [1]. Вдинамически меняющейся среде окружающего мирастратегическими задачами развития системы образованияявляются интеграция науки в учебный процесс в виде научно-исследовательской работы учащихся, развитие междисциплинарных исследований, со стыками в различных областях образовательной деятельности. Все это приводит к необходимости введенияинновационныхметодов и подходов к процессу обучения, в том числе и развитие новых методов проектной деятельности, с целью формирования междисциплинарных знаний и получения практических знаний и умений. Интегративная составляющая проектной деятельности учащихся и индивидуальной траектории развития при личностно-ориентированном подходе достигается за счет применения методов совместной деятельности, что позволяет достичь наилучшего результата.

Вопросы, связанные с проектной деятельностью исследовались Ю.К. Бабанским, А.П. Беляевой, Г.Л. Ильиным, Н.Ф. Талызиной, Т.И. Шамовой, Нагель О.И. и др.В работах [1-5] проведен анализ интегративных процессов в образовательной деятельности. Осуществлен анализ научно-исследовательской деятельности студентов. Представлены пути реализации интеграции учебной и научно-исследовательской работы учащихся через проектную деятельность. Рассмотрены основы, этапы разработки проектов в образовательной деятельности. Показано, что формирование профессиональных компетенций в образовательном процессе является инновационным, достаточно сложным и трудоемким для реализации преподавательским составом вуза, разработаны дидактические модели проектной деятельности, включающее создание благоприятной образовательной среды, позволяющей реально соединить теоретические знания с практическим опытом их применения.

В работах [6-7] учебный процесс показан, как самостоятельный поиск учащимися нового знания, экспериментальное исследование по взаимодействию участников проектной деятельности с применением сетевых коммуникаций при изучении дисциплин программы для формирования общекультурных и профессиональных компетенций студентов вуза.

Исследование совместной деятельности ведется в двух направлениях: совместная деятельность учащегося и педагога и групповая деятельность учащихся. В работе [7] обосновывается актуальность использования групповой работы в учебном процессе, подчеркивается сложность и проблемность применения данной формы организации учебного процесса.

Отмечается, что возможны различные варианты использования групповой работы в учебном процессе: как диагностический, формирование микрогрупп, подготовка обучающихся к групповой деятельности и взаимодействию в микрогруппах, осуществление совместного и индивидуального целеполагания, организация групповой деятельности в коллективе и микрогруппах, представление результатов совместной деятельности, контроль и оценивание групповой деятельности, анализ совместной работы в микрогруппах и учебном коллективе.

В работе [9] выделены следующие характеристики команды: единство цели, совместная деятельность, непротиворечивость интересов, коллективная и взаимная ответственность за результаты совместной деятельности, а также синергетичность взаимодействия членов команды. Выделены «рефлексивные» теоретико-игровые модели, в которых автономность и слаженность совместной деятельности членов команды обеспечивается тем, что их действия согласованы с иерархией взаимных представлений друг о друге.

Однако не всегда возможности известных методов и моделей удовлетворяют потребностям практики и позволяют осуществлять следующие приемы анализа: сравнение полученных результатов личностного развития текущего периода с ранее полученным; оценку реальных тенденций, происходящих в учебном процессе; выявление узких мест и проблем.Возникает необходимость в разработке подходов, позволяющих оценивать траектории развития индивидуальной личности и команды в целом.

Методы оценивания проектной деятельности

Этапы проектной деятельности предполагают планирование задач и действий участников проектов [10], а также получение определенных результатов на каждом конкретном этапе выполнения проекта. Общая схема организации проектной деятельности включает в себя этапы:

1. *Выделение проблемы.* Анализ существующих методов и решений.В ходе работ данного этапа возникает необходимость фиксации проблемы на разных уровнях, а также анализаи подбора спектра готовых технологий, позволяющих сформировать представления о существующих методах решения проблемы. Критерием оценки проектной деятельностиявляется выявление проблемы, требующей междисциплинарного подхода и исследовательского поиска.
2. *Постановка задачи. Формирование гипотезы*. На данном этапе формируется концепция проекта и создается образ конечного продукта. В ходе работ данного этапа задача руководителя проекта состоит в стимулировании обсуждения, консультировании при разработке концепции проекта.
3. *Выбор и реализация проектного продукта*. В этих условиях критериями эффективности могут выступать: организация активной деятельности в учебном процессе, формирование продуктового результата и личностно-ориентированного подхода. Существенную роль при формировании проектной деятельности играет достижение промежуточных результатов работы, к которым в соответствие с жизненным циклом можно отнести: проведение исследования в предметной области, разработка решения проблемы, создание прототипа, испытание и внедрение. Для достижения поставленных целей важной составляющей является работа в группе и управление коммуникацией. Существует множество техник по управлению командой, например сценирование, схематизация, работа внутри команды и взаимодействие с внешним миром. Для внешнего контура проекта целесообразно привлечение экспертов с целью выявления ключевых проблем и применения методологических подходов при моделировании предметной области, а также формирования продуктового результата. Руководитель в этом случае обеспечивает взаимодействие команды с внешним контуром, а также рефлексию внутри команды при взаимодействии с внешним контуром. Для внутреннего контура проекта основные процессы, с которыми работают участники команды, связаны: с мышлением (пониманием того, как устроена модель, какие сценарии работы с информацией необходимо предусмотреть, какие формы взаимодействия возможны при совместном поиске решения задач к кому необходимо обратиться за помощью при необходимости получения дополнительной информации и др.) и с коммуникацией (насколько участник могут взаимодействовать друг с другом, причины конфликтов, использование протоколов взаимодействия, взаимодействие с внешним контуром и др). Это позволяет сформировать

необходимые компетенции по организации коллективного действия и личностные компетенции (совместная деятельность и навыки проектирования) в достижении результата.

4. *Получение практической значимости продуктового результата* и оформление конечных результатов полученного продукта.На этапе реализации проекта возникает необходимость в экспертной дискуссии по валидациии верификации разрабатываемого продукта и формированииобратной связи от пользователей, а также в получении рекомендаций для улучшения проекта. Этап защиты полученных результатов позволяетперейти от рефлексии к взаимодействию с экспертами на стадии запуска продукта, а такжеполучению критериев оценки эффективности работы по проекту.

Построение работы на каждом из этапов проекта требует введениеколичественной оценки качественных показателей результативности как учебного проекта, так и каждого обучаемого по определенным критериям.Предложенный в [11] методпозволяет оценить критериисформированности у студентов профессионально-прикладных компетенций, обеспечивающих образовательный результат. Для этого необходимо:

1. На каждом из этаповс учетом групповой работы студентамиотобрать не менее 5-10 показателей, которые являются важнейшими при определении траектории индивидуального развитияи/или развития проекта.

2. Осуществляется расстановка показателей z_i по рангам с первого, представляющегося наиболее важным, до n-ого, который, по мнению студента наименее значим. Если же несколько показателей, по мнению j-ого студента имеет одинаковое влияние (вес), то последовательные их места суммируются и вычисляется среднее значение, которое и присваивается в качестве ранга данным одинаковым показателям.

3. Оценивается среднее значение места i-ого показателя по данным всех m студентов:

$$z_i = \frac{1}{m}\sum_{j=1}^{m} c_{ij}\,, \tag{1}$$

здесь: c_{ij} - ранг i-ого показателя, назначенный j-ым обучаемым.

4. Рассчитывается среднее квадратичное отклонение i-ого показателя места от его среднего значения z_i:

$$\sigma_i^2 = \frac{1}{(m-1)}\sum_{j=1}^{m}(c_{ij} - z_i)^2 \tag{2}$$

5. Далее осуществляется выставление предварительных рангов по всем i-тым показателям по схеме: если некоторое значение z_i является наименьшим из всех остальных, то ему назначается предварительный ранг $r_i = 1$, далее последовательно выбирается следующий наименьший по величине ранг и ему назначается предварительный ранг 2; если некоторое подмножество средних значений показателей места отличаются не более чем на Δz,то всем им назначаются одинаковые предварительные ранги, как среднее по числу их возможных порядковых мест:

$$\Delta_z = \left|z_{i+1} - z_i\right| \le 0{,}1\left(\frac{n^2}{m-1}\right)^{0{,}25}. \tag{3}$$

Ранжирование позволяет сформировать оценку значимости показателей при обучении студентов и выявить как наиболее приобретение направления в личностном развитии, как студента, так и группы в целом. Это позволяет более тщательно сформировать учебный процесс и осуществить планирование мероприятий для его совершенствования.

6. Определение достижения компетентности может быть выполнена по взаимному оцениванию. Для этого по достижению ключевой точки развития составляются списки студентов в количестве экземпляров по числу m обучающихся; каждый оцениваемый индивидуально проставляет ранги своим сокурсникам, а также же ранговую самооценку; производится расчет коэффициент компетентности по формуле

$$\alpha_j = \frac{1/r_i}{\sum_{j=1}^{m} 1/r_i}, \quad (4)$$

где r_i-значение ранга j оцениваемого по оценкам местстудентов. Это позволяет выявить не только компетентность обучаемого, но и его роль в проекте при групповой работе.

7. Для определение согласия при совместной деятельностистроится матрица вероятности $p_{ik} = \frac{m_{ik}}{m}$, где m_{ik} количество обучаемых, установивших i-му показателю k-ое место по значимости. Коэффициент согласия определяется по формуле:

$$W_H = 1 - \frac{H}{n \log n}, \quad (5)$$

где $H = \sum_{i,k=1}^{n} P_{ik} \log P_{ik}$ Если значение $W_H \geq 0{,}55$ можно считать, что групповая работа состоялась и перейти к следующему этапу.

Результаты оценки проектной деятельности

В экспериментальном исследовании принимали участие две группы студентов пятого курса технического ВУЗа. Для первой группы предусматривалось обучение со сдачей индивидуального курсового проекта в рамках традиционной системы в пределах учебной программы, для второй группы – совместная работа над проектом, в пределахописанного ранее метода. Для анализа результатов, полученных при обучении обоих групп, были выделены главные характеристики оценки траектории развития личности, ранжированные в табл 1. согласно формулам (1-3).

Анализ полученных данных показал, что в пределах групповой работы возникает потребность в развитии навыков коммуникации, однако снижается потребность в коммуникации с наставником. Необходимость получения теоретических навыков также снижается, и это может быть связано как с получением теоретических знаний внутри группы от сокурсников, так и с нацеленностью на результат получения практического опыта. Личностная оценка внутри групп показала, что у лидеров внутри группы с индивидуальными заданиями имеется потребность в развитии навыков работы с наставником, тогда как у группы с проектным заданием –в развитии навыков совместной деятельности.

Таблица.

Ранжирование характеристик траектории развития

Характеристики	Ранг	
	Группа с индивидуальными заданиями	**Группа с проектным заданием**
Практические навыки	1	1
Теоретические основы навыки	2	2,5
Навыки коммуникации	3,5	2,5
Работа с экспертами	5	5
Коммуникация с наставником	3,5	4
Потребность развития проектного опыта	6	6

Оценка компетентности лидера в проектной группе совпала с ожиданиями преподавателя, и оценкой группы в целом, тогда как оценка коэффициента компетентности лидера в группе с индивидуальным заданием имеет существенные расхождения (более 20%) с оценкой, выполненной преподавателем.

Согласованность оценок выше в группе с проектным заданием по сравнению с группой, имевшей индивидуальные задания. Расхождения в группе с проектным заданием имеется в трех характеристиках, тогда как в группе с индивидуальным заданием практически по всем характеристикам.

В результате анализа применения разработанной методики можно предположить, что связь проектной деятельности и методики её оценки может стать критерием сформированности у студентов профессионально-прикладных компетенций, обеспечивающих проектный результат в ходе освоения образовательной программы за счет интегративных процессов, полученных в ходе совместной деятельности.

Заключение

Анализ результатов проведенных исследований по моделированию проектной деятельности среди студентов технического ВУЗа показал следующее. Текущие методы оценивания траектории личностного развития студентов не в полной мере удовлетворяет потребностям практики. Одной из причин такого состояния выступает несовершенство научно-методического аппарата оценивания личностного развития при групповой работе студентов. Предложенный метод по оценке состояния совместной работы позволяет осуществить прогноз развития проекта и осуществить планирование мероприятий для его совершенствования.

Совокупность полученных результатов, к которым приводит проектная деятельность,может рассматриваться, как важное условие для подготовки к профессиональной деятельности и личностного развития студентов, а также отработки практических навыков планирования и управления проектами.

ЛИТЕРАТУРА

1. Тюлю Г. М., Старшинов В. Н. Проектная деятельность как условие интеграции научно-исследовательской и учебной деятельности студентов в образовательном процессе вуза // Вестник Костромского государственного университета. Серия: Педагогика. Психология. Социокинетика. 2016. №22. С. 172-174.
2. Сафонова К. И., Подольский С. В. Проектная деятельность студентов в вузе: принципы отбора проектов и критерии формирования проектных групп // Общество: социология, психология, педагогика. 2017. №9. С. 52-61. DOI: 10.24158/spp.2017.9.11 .
3. Семенова О.Л., Карась С.И., Аржаник М.Б., Острикова О.И., Корнева И.О. Формирование профессиональных компетенций в ходе проектного обучения студентов медицинского вуза // Медицинское образование и профессиональное развитие. 2015. №4 (22). С. 108-121.
4. Как стать наставником проектов: Просветительский проект Лекториум. [Электронный ресурс]. 2018. Дата обновления: 05.08.2018. URL: https://www.lektorium.tv/node/32055 (дата обращения: 05.08.2018).
5. Воробьев В. И., Фаткиева Р. Р. Вопросы обучения в междисциплинарном пространстве // Материалы III межрегиональной научно-практической конференции Перспективные направления развития отечественных информационных технологий / Севастополь, 19-23 сентября 2017 г. С. 99-201
6. Кручинин М.В., Кручинина Г.А. Взаимодействие участников проектной деятельности в вузе с применением сетевых коммуникаций //Казанский педагогический журнал. 2015. № 6 (113), Часть 2. С. 290-296 .
7. Камалеева А.Р., Шигапова Н.В., Шигапов А.А. Научно-методические конструкты построения электронного образовательного ресурса в естественнонаучном образовании // Научное обозрение. Педагогические науки. 2016. №4. – S.16-22
8. Байбородова, Л. В. Организация групповой работы в учебном коллективе / Л. В. Байбородова, С. В. Данданова//Ярославский педагогический вестник. 2016. № 6. С. 75-83.
9. Новиков Д.А. Математические модели формирования и функционирования команд. М.: Издательство физико-математической литературы, 2008. 184 с.
10. Организация проектной деятельности : учебное пособие / Е. В. Михалкина, А. Ю. Никитаева, Н. А. Косолапова. Южный федеральный университет. – Ростов-на-Дону : Издательство Южного федерального университета, 2016. 146 с.
11. Николаев В.И., Фаткиева Р.Р. Системный анализ. Раздел 1. Системное проектирование: Учеб. Пособие. СПб.: СЗТУ, 2002. 141 с.

ВОПРОСЫ ПОДГОТОВКИ IT-СПЕЦИАЛИСТА В УСЛОВИЯХ ЦИФРОВОЙ ТРАНСФОРМАЦИИ

Е**. Е. Котова**[1]
(*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И.Ульянова (Ленина), С.-Петербург, eekotova@gmail.com*)

Аннотация

Данные в цифровой форме становятся ключевым фактором производства во всех сферах социально-экономической деятельности, включая и научно-образовательные организации, что отмечается в Программе «Цифровая экономика Российской федерации» (http://government.ru).

Экспоненциальные темпы технологических изменений приводят к изменениям и в сфере подготовки профессиональных кадров. IT-компетентность является важнейшим показателем уровня подготовки высокопрофессионального специалиста в вузе. Рассматриваются тенденции развития сферы ICT-индустрии (Information and Communication Technology) в контексте подготовки IT-специалистов.

Вопросы организации процесса обучения с применением электронных средств, сочетающие традиционные методы аудиторного обучения с инновационными прогрессивными технологиями требуют новых форм и методов на основе клиентоориентированного подхода и применения принципов CRM-систем. Представлена функциональная структура интегрированной веб-среды сопровождения процесса обучения.

Ключевые слова: интегрированная среда обучения, подготовка IT-специалиста, клиентоориентированный подход.

Введение

Определяющим трендом современных образовательных систем являются цифровые технологии и электронные средства обучения.

К базовым направлениям развития цифровой экономики, определенным в Программе «Цифровая экономика Российской федерации» (http://government.ru)., относятся кадры и образование, формирование исследовательских компетенций. В Программе также отмечается необходимость разработки профилей компетенций специалистов.

Действительные изменения цифровой экономики диктуют новые стратегии построения процесса обучения, ориентированного на создание информационных систем «поддержки индивидуального профиля компетенций обучающегося», внедрение персональных траекторий обучения, «позволяющих обучаемым выбирать индивидуально способы (формальные, неформальные, информальные) формирования базовых компетенций цифровой экономики» с использованием дистанционных образовательных технологий и электронного обучения, что отражено в Программе "Цифровая экономика Российской Федерации".

Интегрированная среда обучения на базе новых технологий может предоставить студентам платформу интерактивного общения и высококачественные электронные ресурсы в дополнение к традиционному режиму обучения. С помощью цифровых ресурсов возможно организовать виртуальные лаборатории, тренажеры, симуляторы, онлайн-моделирование, коллективные взаимодействия с целью формирования исследовательских компетенций.

Цель настоящей работы – рассмотреть вопросы подготовки IT-специалистов и способы организации интегрированной веб-среды сопровождения процесса обучения на основе клиентоориентированного подхода с применением принципов CRM-систем.

Требования к работникам цифровой экономики

Новые рабочие места формируют самые высокие требования к работникам цифровой экономики. Умения генерировать и обрабатывать сложную информацию, идентифицировать и решать реальные проблемы, понимать сложные концепции, связи, иметь развитые когнитивные навыки в работе с информацией, использовать ресурсы с помощью онлайн-инструментов востребованы в новых областях Data-аналитики, Data Science, Data engineering и др.

В перечне приоритетных видов экономической деятельности (по материалам информационно-аналитического отчета «Анализ мирового опыта развития промышленности и подходов к

[1] Кандидат технических наук, доцент каф. АПУ СПбГЭТУ «ЛЭТИ».

цифровой трансформации промышленности государств-членов Евразийского экономического союза», январь 2017, http://www.eurasiancommission.org) предусмотрены «сектора научно-технической и инновационной сферы, которые тесно связаны с цифровой трансформацией промышленности и формированием единого цифрового пространства», среди которых: информационно-коммуникационные и компьютерные технологии, технологии больших данных, суперкомпьютерные технологии, когнитивные технологии, технологии искусственного интеллекта, робототехника, цифровое проектирование и моделирование и др., ориентированные на «создание новых инновационных секторов промышленности» [1]. Там же отмечаются стратегии развития крупнейших компаний: «крупнейшие промышленные и индустриальные компании мира реализуют стратегии развития в концепции «Индустрия 4.0», «Интернет +» (например, Siemens, General Electric, SAP, Intel)». Приводятся данные по цифровым стратегиям, утвержденным в Европейском Союзе – «Цифровая Европа 2020» (2010 г.), «Индустрия 4.0.» (Германия, 2011 г.), «Интернет плюс» (Китай, 2015 г.). Базовые технологии для цифровой трансформации промышленности включают технологии искусственного интеллекта, облачные технологии, большие данные, туманные вычисления, интернет вещей, мобильные технологии, технологии идентификации, математического моделирования, технологии промышленной аналитики (Аналитики 3.0) и др. [1].

Gartner (американская компания, специализирующаяся на исследованиях рынка информационных технологий) прогнозирует, что Artificial Intelligence, (AI), машинное обучение вместе с Интернетом вещей (Internet of Things, IoT) в течение нескольких следующих лет кардинально изменят все рынки. Также Gartner предсказывает, что в 2020 году 95% всей новой электроники будут использовать IoT [2].

Вместе с тем, в связи с политикой импортозамещения многими компаниями поддерживаются отечественные решения российских разработчиков по созданию IT-систем высокого класса, а также решения на базе Open Source.

Тенденции развития сферы ICT-индустрии (Information and Communication Technology), переход от автоматизированных средств к автоматическим и еще большее сокращение рутинных и ручных операций, показывают, что востребованность на рынке труда работников умственного труда (knowledge workers) очевидна. IT-специалисты и ICT-специалисты в области информационных и коммуникационных технологий должны будут владеть больше, чем традиционными ИТ-навыками. Необходимо будет разбираться в «добыче» данных, понимать характер данных, уметь анализировать данные из нескольких источников, разрабатывать модели, создавать метаданные для обеспечения долгосрочного управления данными, прогноза и принятия решений.

По данным Департамента информационных технологий Москвы в 2018 году произойдет взрывной рост платформ, систем и мобильных приложений на базе искусственного интеллекта и нейронных сетей [3].

Большинство применений методов машинного обучения относится к предиктивной аналитике (Predictive analytics) – 26%, к описательной аналитике (Descriptive analytics) – 23%, также данные методы применяются в области управления роботами – 14%. [3].

Новые приоритетные направления и востребованные профессии будущего формируют новые компетенции. Следствием является необходимость модернизации форматов подготовки специалистов для высокотехнологичных отраслей, в частности бакалавров по направлению подготовки 09.03.02 – «Информационные системы и технологии» (профиль «Информационные системы и технологии в бизнесе»), магистров 09.04.02 – «Информационные системы и технологии» (магистерская программа «Управление IT проектами и продуктами»), бакалавров по направлению 27.03.04 – «Управление в технических системах» («Компьютерные интеллектуальные технологии управления в технических системах»), магистров по направлению 27.04.04 – «Управление в технических системах» (магистерская программа «Управление и информационные технологии в технических системах»).

Как отмечает директор Федерального исследовательского центра «Информатика и управление» Российской академии наук академик РАН И.А. Соколов, «сегодня нужны специалисты, которые могут не только пользоваться системами, цифровыми инструментами, но и разрабатывать соответствующий инструментарий…. Во многих университетах мира ведущие позиции

занимают науки, основанные на интенсивном использовании данных. Исследования в данном направлении обеспечивают возможность собирать данные о явлениях и объектах, обрабатывать их новыми методами, чтобы извлекать знания. Важно разработать стандарты подготовки таких специалистов» [4]. Академик отмечает, что «уже с 2010 года в ведущих университетах мира появились курсы Big Data, появились компетенции, неожиданные для нас, которые требуют определенных специалистов. Стандарты в области информационных технологий быстро устаревают, а технология подготовки загнана в такие рамки, что их изменение – чрезвычайно долговременный и трудоемкий процесс. Так или иначе, проблему подготовки кадров, обладающих актуальными компетенциями, нужно решать… В МГУ создан Национальный центр цифровой экономики для исследований и подготовки кадров. Его задача, в том числе, разрабатывать методические рекомендации и учебные материалы для того, чтобы корректировать существующие, либо создавать новые учебные программы» [5].

Клиентоориентированный подход к подготовке IT-специалистов. Принципы CRM-систем.

Клиентоориентированный подход является ядром концепции CRM (Customer Relationship Management) – концепции управления взаимоотношениями с клиентами, направленной на применение новых управленческих и информационных технологий. В отличие от клиентов компаний и предприятий «клиентами» университетов являются обучающиеся. Принципы CRM-систем в сфере образования применяются во многих университетах. Внедрение концепции CRM в высшие учебные заведения стало реальностью многочисленных учреждений по всему миру [6 – 10]. Исследования по применению принципов маркетинга бизнес-сектора к контексту высшего образования начали появляться в 1990-х годах [7]. Эмпирические и теоретические работы, направленные на выявление ключевых факторов применения CRM систем в образовании, опирались на использование теории, разработанной и хорошо зарекомендовавшей себя в области управления бизнесом. Основное внимание уделялось деловой терминологии, вопросам, могут ли студенты считаться «клиентами», что является «продуктом» и «услугой» высшего образования [7]. Позднее было признано, что высшее образование не является продуктом, а является услугой, и что маркетинг услуг в значительной степени отличается от маркетинга продуктов. Ключевой характеристикой, обеспечивающей основу для маркетинга услуг, определялся характер услуги, то есть подчеркивалась важность отношений с клиентами [7]. По результатам ряда исследований [11] также отмечаются преимущества модели маркетинговых отношений для некоммерческих организаций, таких как университеты.

Основной функцией CRM-системы в образовательной организации является поддержка и развитие отношений с абитуриентами, студентами и партнерами. Авторы по аналогии с CRM вводят аббревиатуру SRM – Student Relationship Management в контексте управления в высшем образовании [10].

CRM обычно ассоциируется с бизнес-контекстами. Однако, в [12] отмечается, что принципы и приложения CRM также подходят для некоммерческих организаций. Сфера применения CRM из бизнес-области может быть расширена и иметь значительный потенциал в области образовательных учреждений. По мнению авторов [12], этот вопрос имеет большой потенциал, но пока исследованиям уделяется мало внимания. Авторы используют термин «e-Monitoring Systems» для отражения эффективности отношений в электронном обучении, которое постепенно вводится в традиционные процессы обучения.

Конечная цель, которую преследует любая организация, использующая CRM-системы – это предоставление своим клиентам дополнительных возможностей на основе понимания их индивидуализированных предпочтений.

Клиентоориентированный подход подразумевает предпочтение индивидуальной работы со студентом и предоставление учебного материала в зависимости от его предпочтений, целей и индивидуального темпа восприятия информации. Клиентоориентированный подход предполагает глубокое понимание потребностей клиента и их эффективное удовлетворение.

Цель клиентоориентированного подхода в предложении студентам полного комплекта услуг в соответствии с программами подготовки, но в то же время с учетом индивидуальных предпо-

чтений профессиональной направленности, а именно прохождения бакалавриата, магистратуры, возможно аспирантуры с дальнейшим трудоустройством. Сопровождение развития учащегося в динамике – от абитуриента, затем студента, и до выпускника является приоритетной задачей клиентоориентированного подхода.

По мнению исследователей, в основе реализации клиентоориентированного подхода лежат механизмы информационного обеспечения на основе поддержки клиентской базы с применением принципов CRM-систем [13]. CRM-системы реализуют функции автоматизации сбора, обработки, анализа и визуализации маркетинговой информации, которая позволяет получить детализированную структуру потребностей потенциальных клиентов [13].

Задачи, выполняемые системами CRM в высшем образовании, делятся на две основные области – операционные и аналитические. Операционная система собирает данные от различных источников; аналитическая система анализируют эти данные [8]. При внедрении CRM-системы необходимо решить, какая информация нужна о клиенте, как информация собирается, где хранятся данные, как и кем она используется [8]. Используя данные, полностью представляющие каждого клиента, аналитическая система должна дать точное указание, где и в какой форме необходимы дополнительные услуги и углубленное консультирование студентов, необходимо понимание каждой «уникальной» ситуации учащегося [8]. Большее число поступающих в вузы имеют значительный опыт работы с информационными технологиями. Этот опыт говорит о высоких ожиданиях студентов относительно доступных технологических ресурсов.

Учащиеся строят свои познавательные стратегии в зависимости от уровня предыдущих знаний и учебной среды. Для развития личностный и профессиональных компетенций требуется богатая и стимулирующая среда обучения. Необходима разработка подходящей среды, т. н. «глубокой среды обучения» [14], ориентированной на сотрудничество и индивидуальные предпочтения обучающихся разного уровня подготовки, как начинающих, так и более успешных.

В нашей работе предложена интегрированная веб-среда сопровождения процесса обучения, функциональная структура изображена на рис. 1. Веб-среда предполагает интеграцию традиционных технологий обучения с электронными, что в настоящее время относится к классу систем смешанного, интегрированного, гибридного обучения (B-Learning, Blended Learning). С развитием интернет-технологий возможно реализовать функции CRM-решений на основе онлайн-методов управления в веб-среде с применением агентного подхода.

Интегрированная веб-среда сопровождения процесса обучения

Программные интеллектуальные агенты на основе использования аналитических инструментов способны собирать, систематизировать необходимую информацию, осуществлять анализ данных о процессе обучения, выявлять ключевые показатели, прогнозировать результаты. Поведение агентов запрограммировано и регулируется агентом-менеджером. В системе предусмотрены агенты: преподавателя (Агент *TUI*, Teacher User Interface), аналитика (Агент *Anl*, Analyst), студента (Агент *SUI*, Student User Interface) и другие (агенты сбора, подготовки, анализа данных, построения моделей, идентификации параметров студентов, прогнозирования результатов обучения, управления распределением учебной когнитивной нагрузки, построением сценариев обучения в персональной среде). Агенты взаимодействуют посредством языка agent communication languages (ACL).

С помощью агентов поддерживается следующая функциональность:

- анализ целевой аудитории;
- «обслуживание студентов» (Student Service) – предоставление дидактических ресурсов, сопровождение процесса обучения, академическая поддержка, консультирование по вопросам обучения и карьеры;
- мониторинг (Monitoring);
- отчетность (Reporting);
- аналитика (Analytics);
- прогноз (Prediction).

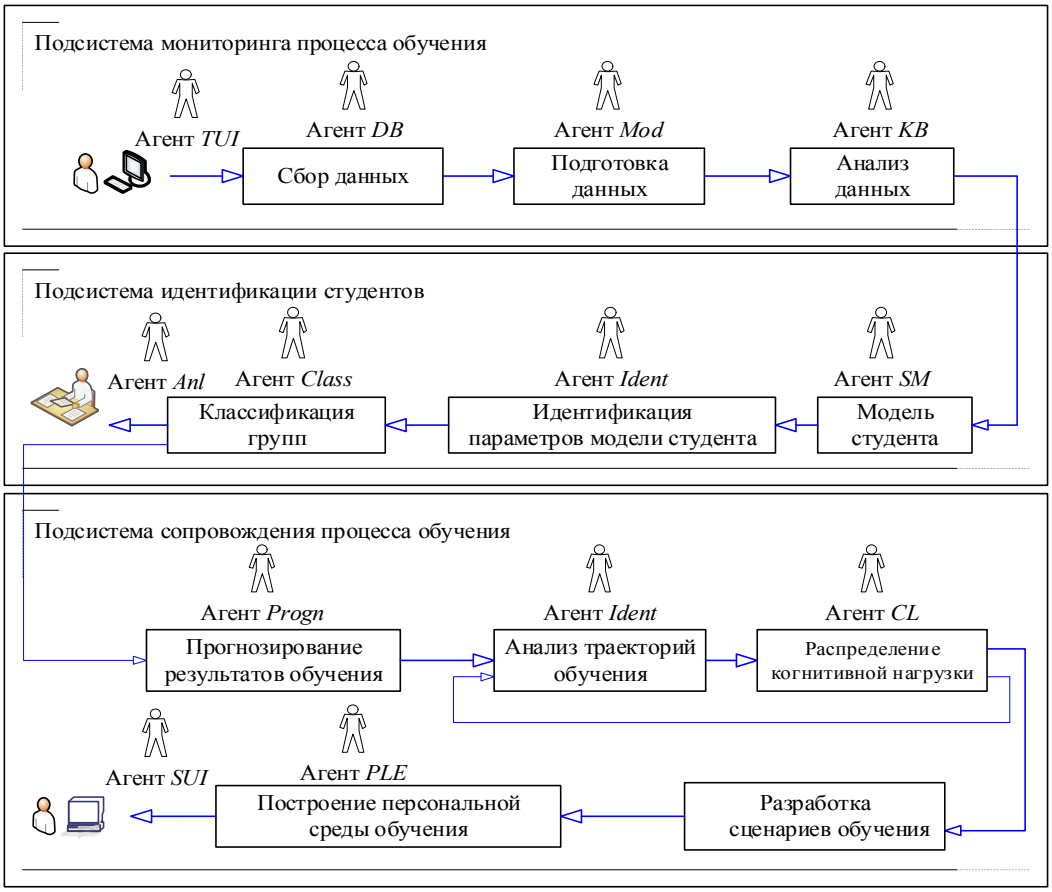


Рис. 1. Функциональная структура интегрированной веб-среды сопровождения процесса обучения

Сбор данных - это первый шаг в идентификации данных, которые могут выявить полезную и ценную информацию среди объемов данных. После того, как данные были отобраны и подготовлены для использования, необходим их анализ. Однако с ростом разнообразия характера данных, анализ данных становится очень сложным процессом. Анализ должен включать сопоставление различных наборов данных для понимания информации, которая должна быть доступна для пользователя, интерпретироваться и интегрироваться в существующие процессы. Данные, представляющие объемы информации о студентах, включая демографические, академические и др. записи, необходимы для использования целевой аналитики. Методы интеллектуального анализа образовательных данных и аналитики (Educational data mining, EDM) в высшем образовании могут изменять существующие процессы администрирования, преподавания, обучения, академической работы, способствуя решению современных проблем, стоящих перед образованием. Большие данные могут предоставить высшим учебным заведениям прогностические инструменты, необходимые для улучшения результатов обучения отдельных учащихся, а также способы обеспечения академических программ высококачественными материалами.

Первоочередной задачей является анализ целевой аудитории. Агент базы данных (агент *DB*) собирает всю необходимую информацию о студенте. Одно из основных преимуществ в результате реализации SRM - это рассмотрение интересов студентов [10]. Первым шагом может быть выявление интересов через мотивацию студентов. Мотивация является основным фактором успеха любой учебной среды. Источник мотивация варьируется от внешнего (вне учебной среды) до внутреннего (часть учебной среды). Мотивацию учащегося трудно идентифицировать, но это необходимо учитывать при разработке [14].

На рис. 2 в качестве примера приведены данные по исследованию мотивации группы первокурсников (48 человек – студентов факультета ФКТИ СПбГЭТУ «ЛЭТИ» по направлениям подготовки «Информационные системы и технологии» и «Компьютерные интеллектуальные технологии управления в технических системах»). Показатели, которые были выявлены после заполнения мотивационной анкеты, по которым проводился дальнейший анализ, перечислены в таблице 1.



Рис. 2. Результаты исследования мотивации студентов

В приведенном небольшом исследовании мотивации студенты проявили интерес к дисциплинам профессиональной направленности (42 %), а также заинтересованность в перспективах, в будущем трудоустройстве уже с первого курса (18%). Отмечена значимость репутации университета, традиции и положение в рейтинге (38%), опыт и квалификация преподавателей (15 %). Местоположение университета (город) является важным критерием для 44% абитуриентов. Знакомые, друзья и родственники были признаны важными вторичными группами лиц, принимающими решения для выбора места обучения (27 % студентов). Привлекает атмосфера в вузе (16%), отмечена архитектура зданий (15%). Тем не менее, для 42 % выпускников средней школы прежде всего важно, что предлагаемые программы курсов соответствуют их специализированным интересам к информационным технологиям и программированию (42 %).

Таблица 1

Показатели мотивации

№	Показатель	% студентов
1	Местоположение вуза (город Санкт-Петербург)	44
2	Интерес к информатике, программированию, IT-технологиям	42
3	Репутация вуза, традиции, высокий рейтинг	38
4	Советы знакомых, друзей, родственников	27
5	Перспектива трудоустройства, карьерного роста	18
6	Атмосфера в вузе	16
7	Авторитет, опыт, квалификация преподавателей	15
8	Архитектура вуза	15

Агент базы данных обеспечивает сбор информации и отчетность как для аналитических целей, выполняемых агентом аналитики, таких как интеллектуальный анализ данных, так и для оперативных задач, выполняемых агентами анализа траекторий обучения и прогнозирования.

Одним из применения методов интеллектуального анализа данных является идентификация типологии студентов, которая используется для разделения студентов на разные группы обучения. Идентификация студентов требует большого внимания для построения индивидуального обслуживания [9]. Кроме того, одна из задач идентификации студентов – сокращение числа «слабых» студентов (предотвращение «отсева»). Могут быть сгенерированы прогнозы успешного завершения обучения, и, в соответствии с этим, скорректированы программы индивидуального обучения студентов. Аналогичные подходы используются, например, в [15], где отмечается, что результаты интеллектуального анализа данных могут быть использованы для лучшего распределения ресурсов, активного управления результатами учебы и повышения эффективности развития выпускников в учебных заведениях [15]. Как отмечает автор [15], выявление соответствующих типологий, вновь обнаруженные закономерности помогают педагогам и ад-

министраторам лучше удовлетворять потребности различных групп студентов. Классификация и прогнозирование – два основных метода, которые необходимы для обнаружения образа успешного/неуспешного студента.

Способы усвоения учебного материала, как и способы подачи его педагогами претерпевают значительные изменения, что «отчасти является результатом нового понимания процесса обучения, а отчасти – результатом новых технологий» [16]. При разработке программ с использованием данных об учащихся, с помощью оперативной обратной связи, необходимо предусмотреть в учебной программе индивидуальные модули и задания.

Ввиду развития цифровых технологий, которые требуют развитых когнитивных навыков работы с информацией в составе IT-компетентности, начиная с первого курса, например, в дисциплинах «Информатика», «Программирование», целесообразно включать в практические занятия и лабораторные работы задания по самостоятельному анализу данных и обмену ими, средствам анализа и визуализации данных и знаний.

Например, практические задания по «Информатике» в 1 семестре для студентов 1 курса направления подготовки бакалавров 09.03.02 – «Информационные системы и технологии» (профиль «Информационные системы и технологии в бизнесе») уже могут включать элементы изучения бизнес-процессов. Далее приведем примеры индивидуальных практических заданий.

Пример1. Практическое задание по дисциплине «Информатика» в рамках направления развития управления бизнес-процессами ACM – Adaptive Case Management, адаптивного ведения дел: сформировать контрольный список действий для получения итогового зачета по дисциплинам 1 семестра и представить в виде, например, чек-листа. Понять, что такое чек-лист не представится сложным для студента-первокурсника. При этом во время выполнения задания появится ориентировка в процессе обучения (по нашим данным большая часть студентов 1 курса в начале первого семестра слабо ориентируются в структуре учебного процесса и в системе контроля знаний и получения результатов). Выполняя данную практическую работу студент достаточно легко разберется, что от него требуется для получения зачетов, и сам определит наиболее продуктивный стиль работы (учебной деятельности). И, что тоже важно, всегда сможет быстро себя перепроверить, если чек-лист создан средствами простой таблицы, например, Microsoft Excel. Можно представить такую работу как микроуровень организации бизнес-процесса. В коллективном обсуждении выявить наиболее интересные аспекты, связанные с изменчивостью «бизнес-процессов», изменчивостью их «составных элементов». По результатам анализа определить влияющие факторы на процесс. По нашему опыту, аналогичные задания повышают мотивацию у студентов к обучению.

Пример 2. Практическое задание по дисциплине «Информатика» на структурирование и визуальное представление областей знаний с применением методов извлечения знаний. Задачей является поиск информации из неструктурированного объема данных по заданной теме. Актуальность такой задачи подтверждается, например, словами академика И.А. Соколова: «Извлечение знаний из неструктурированного объема данных, например, из речи, текста, – задача искусственного интеллекта. В науке это, возможно, одна из наиболее приоритетных задач информатики и математики, решение которой требуется в целях цифровой трансформации» [4]. Этап извлечения знаний является начальным этапом процесса интеллектуального анализа и обработки знаний. Начальные навыки методов извлечения знаний являются предпосылкой к освоению технологий извлечения знаний, которые входят в программы дальнейших дисциплин: инструментальных средств запросов, статистических методов, методов визуализации, интерактивной аналитической обработки (OLAP), обучения, основанного на прецедентах (k-ближайший сосед), деревьев решений, ассоциативных правил, нейронных сетей, генетических алгоритмов и др..

Различные по сложности индивидуальные задания были предложены студентам, рекомендации заданий формировались персональным агентом когнитивной нагрузки в зависимости от уровня знаний.

Один из подходов маркетинговых стратегий, которые используют университетские учреждения – подход удовлетворенности процессом обучения студентов [7].

Университеты во всем мире признают необходимость изменения их бизнеса в направлении развития наилучших отношений со студентами [9]. Управление взаимоотношениями с клиен-

тами – это и бизнес-стратегии, и набор программных инструментов и технологий с целью выявления новых возможностей для увеличения удовлетворенность клиентов (студентов) [9].

На рис. 3 приведен пример результата опроса студентов двух групп удовлетворенностью процесса обучения с применением разных форм сопровождения обучающихся – с помощью сопровождения процесса обучения в онлайн-среде с применением персональных агентов и без сопровождения. Выборку составили студенты двух групп (в каждой группе по 24 студента) первого курса.

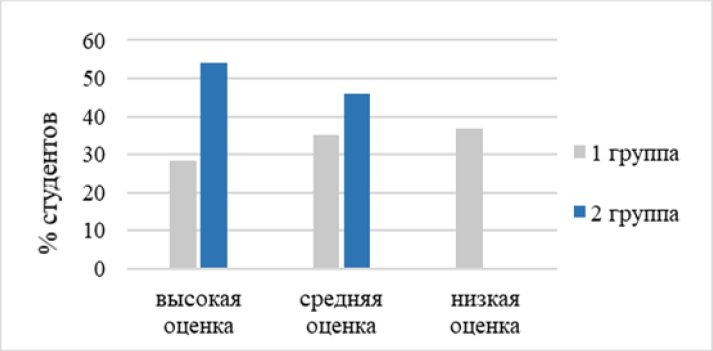


Рис. 3. Результаты опроса (сравнение двух групп)

Группа 2 обучалась с системой онлайн-поддержки, которая дополняет традиционную учебную систему, существующую в настоящее время. В среде использовались индивидуальные формы заданий с учетом уровня подготовки студента и на основе предварительной диагностики индивидуальных когнитивных параметров восприятия информации [17].

Результаты демонстрируют значимые различия в оценках двух групп – большую удовлетворенность процессом обучения студентами второй группы.

Заключение

В докладе рассматриваются вопросы подготовки IT-специалистов в условиях высшего образования, как преимущества реализации клиентоориентированного подхода и принципов CRM-систем. Преимущества внедрения CRM в условиях высшего образования включают управление процессами, ориентированными на студентов, что способствует повышению эффективности учебного процесса.

Ограничения исследований проявляются в объективной сложности реализации всех функций CRM-системы. Однако, на данном этапе организации информационного центра развития цифровых компетенций в структуре научной группы, занимающейся исследованиями в области инженерии знаний (Knowledge Engineering research, ontomaster.ru) на кафедре автоматики и процессов управления СПбГЭТУ «ЛЭТИ», считаем целесообразным организацию внутренней CRM-системы, объединяющей данные из разных источников и баз данных вуза (методический отдел, деканат, преподаватели) на базе виртуальной образовательной среды.

Оригинальность/ценность виртуальной образовательной среды заключается в поддержке E-мониторинга процесса обучения, создании базы студентов, анализе динамики результатов, подборе дидактических ресурсов, управлении индивидуальным развитием компетенций, а также поддержке актуальной информации от работодателей, и др.

Учитывая большое число студентов, совершенно очевидно, что персонализированные услуги представляют собой проблему для учебных заведений. Один из способов решения данной проблемы означает удовлетворение потребностей в услугах большого количества зачисленных студентов средствами внедрения и использование информационных технологий. Достижения в области технологий позволяют предоставить учащимся разнообразные инструменты, как для добавления материала в окружающую среду, так и для оценивания себя – способность учащегося определять цели, планировать и понимать стратегии обучения, что относится к метакогнитивным навыкам, особенно важным при формировании профессиональных компетенций IT-специалиста.

Решение видится в предложении интегрированной платформы обслуживания, которая позволяет персонализировать доступ в соответствии с потребностями студентов. Возможности использования методов интеллектуального анализа данных и управления процессом обучения обеспечивают потенциал для дальнейших исследований

ЛИТЕРАТУРА

1. Информационно-аналитический отчет «Анализ мирового опыта развития промышленности и подходов к цифровой трансформации промышленности государств-членов Евразийского экономического союза». Москва, январь 2017 г. [Электронный ресурс] URL: http://www.eurasiancommission.org
2. 2017: Стимул для замены мейнфреймов. [Электронный ресурс] URL: http://www.tadviser.ru/index.php /Статья:Мейнфрейм_%28мэйнфрейм%29
3. Искусственный интеллект (рынок России). [Электронный ресурс] URL: http://www.tadviser.ru
4. Цифровая трансформация требует компетенций и квалификации. Пресс-конференция «Цифровая трансформация России: современный этап». [Электронный ресурс] URL: http://www.connect-wit.ru/tsifrovaya-transformatsiya-trebuet-kompetentsij-i-kvalifikatsii.html
5. Системный администратор. Выпуск №11 (180). 2017 г. [Электронный ресурс] URL: http://samag.ru/archive/article/3545
6. Философия CRM для образовательных учреждений. [Электронный ресурс] URL: http://www.crm-practice.ru/articles/3227/
7. Hemsley-Brown J., Oplatka I. Universities in a competitive global marketplace: A systematic review of the literature on higher education marketing //International Journal of public sector management. 2006. vol. 19. no. 4. pp. 316-338.
8. Seeman E. D., O'Hara M. Customer relationship management in higher education: Using information systems to improve the student-school relationship //Campus-Wide Information Systems. 2006. vol. 23. no. 1. pp. 24-34.
9. Rodić V., Lukić N., Rodić J. Costumer relationship management in institutions of higher education //II International Symposium Engineering Management and Competitiveness 2012 (EMC 2012) June 22-23, 2012, Zrenjanin, Serbia. 2012. pp. 299-303.
10. Hilbert A., Schönbrunn K., Schmode S. Student relationship management in Germany–foundations and opportunities //Management Revue. 2007. pp. 204-219.
11. Arnett D. B., German S. D., Hunt S. D. The identity salience model of relationship marketing success: The case of nonprofit marketing //Journal of marketing. 2003. vol. 67. no. 2. pp. 89-105.
12. Daradoumis T. et al. CRM applied to higher education: Developing an e-monitoring system to improve relationships in e-learning environments //International Journal of Services Technology and Management. 2010. vol. 14. .no. 1. pp. 103-125.
13. Тарасова К. В. Формирование маркетинговой информационной системы на основе клиентоориентированного подхода//Вестник Белгородского университета кооперации, экономики и права. 2016. №. 3. С. 257-269.
14. Neville K., Adam F., McCormack C. Mentoring distance learners: an action research study //ECIS 2002 Proceedings. 2002. P. 56.
15. Luan J. Data mining and its applications in higher education //New directions for institutional research. 2002. vol. 2002. no. 113. pp. 17-36.
16. Новиков А.М., Новиков Д.А. Методология. М.: СИНТЕГ. 2007. 668 с.
17. Котова Е.Е., Падерно П.И. Экспресс–диагностика когнитивно–стилевого потенциала обучающихся в интегрированной образовательной среде. // Образовательные технологии и общество. 2015. Том 18. № 1.

E.E.Kotova (Saint Petersburg Electrotechnical University "LETI", St. Petersburg)

Issues of training IT-specialist in conditions of digital transformation

The data in digital form become the key factor of production in all spheres of social and economic activity, including scientific and educational organizations, which is noted in the program "Digital Economy of the Russian Federation" (http://government.ru).

The exponential rate of technological change leads to changes in the training of professional personnel. IT competences are the most important indicator of the level of training of a highly professional specialist in a university. The tendencies of development of the ICT industry sphere (Information and Communication Technology) in the context of IT specialists training are considered.

The issues of organizing the learning process using electronic means, combining traditional methods of classroom teaching with innovative progressive technologies require new forms and methods on the basis of a client-oriented approach and application of the principles of CRM-systems. The functional structure of the integrated web environment for tracking the learning process is present.

ТЕХНОЛОГИИ АВТОМАТИЗИРОВАННОЙ ОБРАБОТКИ И ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА НЕСТРУКТУРИРОВАННЫХ ДАННЫХ

Е. Е. Котова[1], **А. С. Писарев**[2], **И. А. Писарев**[3]
(*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина), С.-Петербург, eekotova@gmail.com, a_pisarev@mail.ru, pisarevivan@yandex.ru*)

Аннотация

В работе представлены информационные технологии автоматизации сценариев интеллектуального анализа данных для применения в научных исследованиях и обучении студентов. Разработаны гибридная база данных информационных ресурсов (документов, изображений, видео и др.), база метаданных методов и моделей, средства обработки и интеллектуального анализа данных в Web-среде. Метаданные описывают более 50 методов машинного обучения, относящихся к задачам классификации информационных ресурсов на основе теоремы Байеса (Bayes), деревьев решений (Decision Trees), опорных векторов (Support Vector Machines - SVM), конволюционных нейронных сетей (Convolutional Neural Networks – CNN) и др. Приведены примеры сценариев научных исследований. Автоматизированный анализ данных реализован в сетевой программной среде ОнтоМАСТЕР-Ресурс на основе стандартов семантического Web.

Ключевые слова: автоматизированная система научных исследований, машинное обучение, Web-среда

Введение

Извлечение знаний из неструктурированных данных большого объема относится к одной из наиболее приоритетных задач, решение которой требуется в целях цифровой трансформации экономики. В современных условиях в программы обучения бакалавров и магистров необходимо включать новые методы и технологии извлечения знаний из больших объемов данных (Data Science) [1, 2].

Среди наиболее известных средств для автоматизированного анализа данных применяются программные продукты и платформы для анализа больших данных: WEKA [3], RapidMiner [4], KNIME [5], KEEL [6], SAP Predictive Analytics, Matlab (MathSoft), TenzorFlow [7] и др.

Актуальной задачей является разработка методов и технологий интегрированной обработки и интеллектуального анализа данных в сетевой среде на основе стандартов семантического Web.

Разработка новых средств автоматизации анализа данных относится к направлению исследований в области автоматизации решения задач машинного обучения (Automated Machine Learning – AutoML), целью которого является анализ и извлечение новых знаний из больших объемов накапливаемой информации и данных в разных областях исследований.

Доклад посвящен технологиям автоматизированной обработки и интеллектуального анализа неструктурированных данных (текстов, видео и изображений, сигналов), реализованных в Web-среде и основанным на создании гибридной базы данных и знаний, онтологий методов и процессов обработки, многоагентном подходе с целью выполнения комплексных сценариев научных исследований и сценариев обработки знаний и данных.

Технологии автоматизации интеллектуального анализа данных

Метод автоматизации обработки и интеллектуального анализа данных включает планирование и выполнение сценариев решения задач многоклассовой классификации и регрессионного анализа [8].

Рассматриваются прямая и обратная задачи.

Прямая задача предсказания выходной переменной (класса) от входных переменных (признаков) представлена в виде отображения:

[1] Кандидат технических наук, доцент.
[2] Кандидат технических наук, доцент.
[3] Аспирант.

$$f : X \to Y$$

где f – модель анализа данных.

В случае, если области определения $D(X)$ и изменения $D(Y)$ имеют типы данных «numeric», решается прямая задача регрессионного анализа. Если $D(Y)$ имеет тип данных «symbolic», то решается задача классификации на основании признаков, содержащихся в данных.

Поиск моделей классификации и регрессии формулируется в виде обратной задачи: найти отображение f^* (структуру модели и значения параметров настройки Z) при известных значениях независимых переменных (признаков) X и зависимой переменной (класса) Y, которое обеспечивает экстремум функционала в виде аддитивной свертки целевых функций R_j при выполнении ограничений $R_i \in C$:

$$f^* = \arg\left(extr_{R_i \in C} \left\{ \sum_j \alpha_j R_j \left(f\left(X, Z\right), Y \right) \right\} \right),$$

где α_j – весовой коэффициент критерия R_j, $\sum_j \alpha_j = 1$,

R_j – количественные показатели производительности методов классификации и регрессии.

Качество модели оценивается:

– по одному важнейшему критерию – при этом вес α_j этого критерия равен 1;

– по одному критерию при проверке условия выполнения требований на ограничения для других критериев;

– на основе многокритериальной оценки.

Показателями классификации и регрессии являются:

– среднее значение абсолютной ошибки (MAE, mean absolute error);

– среднеквадратическая ошибка (RMSE, root mean square error);

– относительная абсолютная ошибка (RAE, relative absolute error);

– относительная среднеквадратическая ошибка (RRSE, root relative squared error).

При классификации дополнительно применяются показатели [9]:

– точность (accuracy) – отношение «правильно» (true positive) классифицированных случаев к общему их числу;

– коэффициент каппа-статистики (kappa statistic).

При регрессионном анализе дополнительно используется коэффициент корреляции (CC, correlation coefficient).

Число критериев качества решения прямых и обратных задач целесообразно дополнить также показателями затрат вычислительных ресурсов (памяти, времени) и оценкой возможности «интерпретации человеком» результатов. Свойство «интерпретируемости» модели позволяет реализовать подсистему «объяснения» результатов моделирования в системах поддержки решений.

Реализован поиск моделей и методов, релевантных данным с автоматическим выполнением программ классификации и регрессионного анализа, сохранением результатов в базе данных, формированием отчетов и ранжированием моделей в соответствии с выбранными весами для частных критериев.

В публикациях по применению методов классификации среди наиболее часто используемых критериев отмечается точность метода, а при оценке методов регрессионного анализа – среднеквадратическая ошибка (RMSE) [10, 11].

Предусмотрено вычисление модифицированного информационного критерия Акаике (AIC) что позволяет при выборе лучших моделей учитывать не только перечисленные выше критерии, но и сложность модели в зависимости от числа используемых входных переменных (признаков). В соответствии с принципом Оккама по критерию AICc более предпочтительными являются точные и простые модели [1], [8].

Особенностью подхода является возможность автоматического формирования ранжированного перечня методов и моделей в виде таблицы в порядке убывания одного или нескольких критериев показателей качества моделей.

При оценке качества моделей с помощью модифицированного информационного критерия Акаике учитывается сложность (число параметров) и точность описания экспериментальных данных [12]:

$$AIC_c = 2k + n\ln\frac{RSS}{n} + \frac{2k(k+1)}{n-k-1},$$

где k – число параметров модели, n – размер выборки.

Критерий суммы квадратов остатков (Residual Sum of Squares – RSS):

$$RSS = \sum_{i=1}^{n}\left(y_i^{\text{model}} - y_i^{data}\right)^2,$$

где $y_i^{\text{model}}, y_i^{data}$ – модельные и экспериментальные данные в моменты времени t_i .

Среднеквадратическая ошибка вычисляется:

$$RMSE = \sqrt{\frac{RSS}{n}}.$$

При использовании полученной в результате классификации оценки RMSE, формула вычисления AIC принимает следующий вид для случаев $\text{RMSE} > 0$ и $n - k > 1$ (при 100% точности AIC присваивается условное большое отрицательное число):

$$AIC_c = 2k + n\ln \text{RMSE}^2 + \frac{2k(k+1)}{n-k-1}.$$

Рассматриваются технологии автоматизированной обработки и интеллектуального анализа данных.

Архитектура автоматизированной системы

Архитектура автоматизированной системы состоит из Web-интерфейсов исследователей, программных агентов и сервисов, гибридной базы данных, базы знаний, онтологии методов и моделей [13].

Взаимодействие с автоматизированной системой осуществляется с использованием Web-интерфейсов исследователей и прикладных программных интерфейсов с Web-сервисами и интеллектуальными агентами.

Исследователям предоставляется возможность создания сценариев обработки данных, формирования и тестирования моделей классификации и регрессионного анализа.

Сформированные модели в дальнейшем применяются в автоматизированных информационных системах при решении ряда прикладных задач, примеры которых будут рассмотрены далее. Интеллектуальные программные агенты осуществляют решение задач в соответствии с разработанными сценариями с использованием операций обработки данных, текстовой информации, видео и изображений, сигналов и их интеллектуального анализа.

Гибридная база данных содержит информационные ресурсы (документы, изображения, видео и др.). База знаний включает метаданные о средствах обработки и методах интеллектуального анализа данных. Метаданные описывают более 50 методов машинного обучения, относящихся к задачам классификации информационных ресурсов на основе теоремы Байеса (Bayes), деревьев решений (Decision Trees), опорных векторов (Support Vector Machines - SVM), конволюционных нейронных сетей (Convolutional Neural Networks – CNN), TenzorFlow и др. [3], [7].

Онтологии включают формальное описание в Web-редакторе ОнтоМАСТЕР-Онтология терминов, моделей, методов классификации и регрессионного анализа различных типов объектов. На рис. 1 изображен фрагмент разработанной авторами доклада онтологии методов анализа данных. (рис.1) [14].

Комплексная обработка и интеллектуальный анализ неструктурированных данных осуществляются на основе разрабатываемых сценариев в автоматизированной системе решения задач машинного обучения (Automated Machine Learning – AutoML).

Технологии автоматизированной обработки, классификации и регрессионного анализа данных реализованы в Web-среде ОнтоМАСТЕР–Ресурс [8] и отличаются построением сценариев и выбором методов машинного обучения, релевантных переменным (признакам и классам) данных, что позволяет повысить производительность научных исследований.

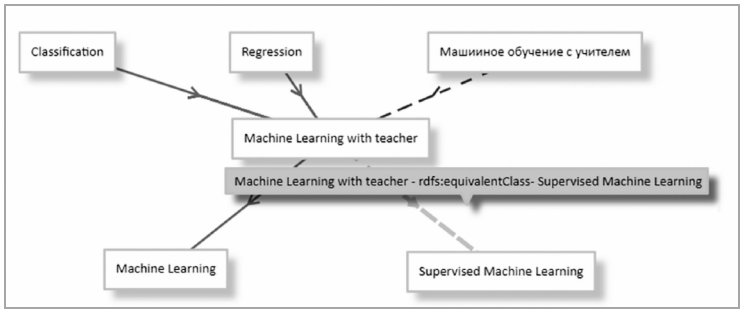


Рис. 1. Фрагмент онтологии методов анализа данных

Программная среда ОнтоМАСТЕР-Ресурс предоставляет возможность визуального редактирования сценариев, раскрытия блоков и назначения онтологических имен классов, реализующих заданные операции. На рис. 2 приведен пример визуального построения алгоритма сценария обработки и анализа данных в Web-редакторе. Сетевая архитектура ОнтоМАСТЕР-Ресурс позволяет проводить совместные научные исследования с использованием Web-интерфейсов и выполнения сценариев по анализу разнородных экспериментальных данных (текстов, видео и изображений, сигналов). Результаты автоматизированного выполнения сценариев сохраняются в базе данных, что позволяет проводить всесторонний анализ и поиск наилучших моделей на основе многокритериального подхода.

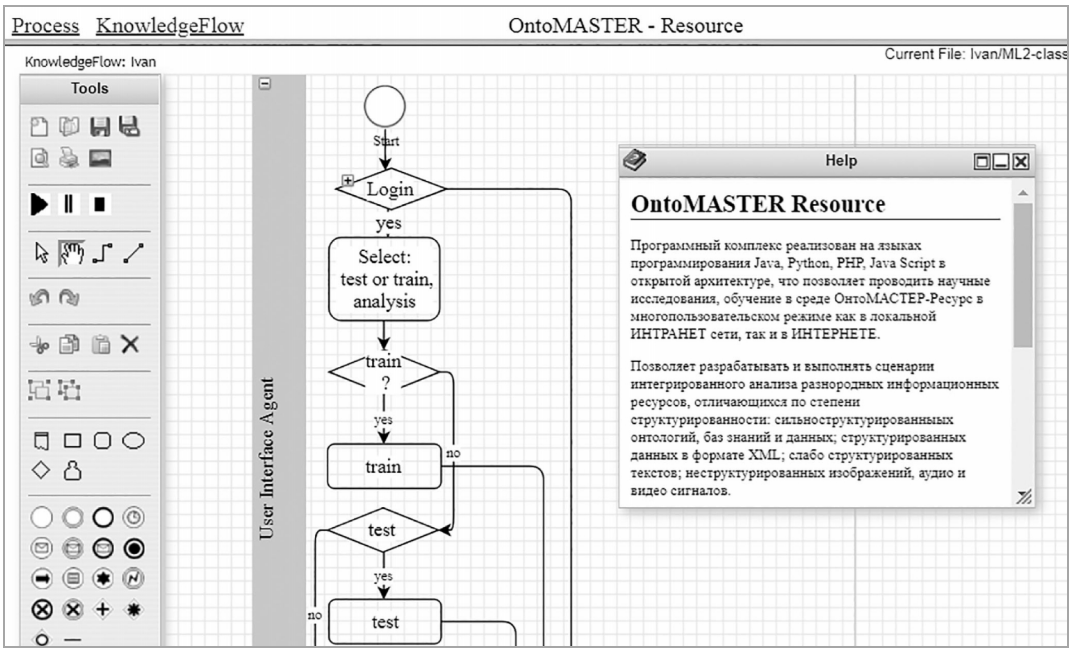


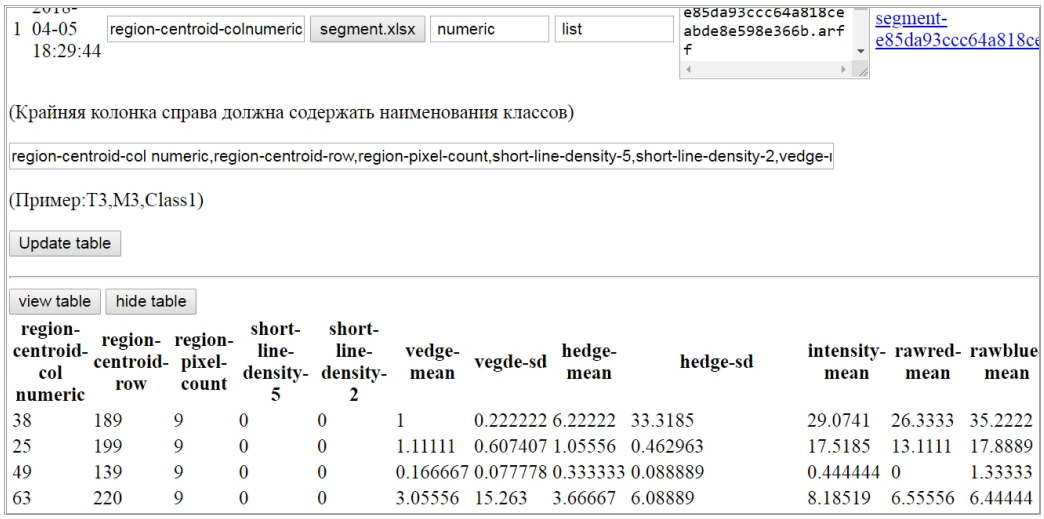
Рис. 2. Web-редактор сценариев обработки и анализа данных

Примеры задач классификации и регрессионного анализа

Тестирование метода и технологий, реализованных в Web-среде ОнтоМАСТЕР-Ресурс, производилось на нескольких задачах классификации и регрессионного анализа.

Задача 1. Классификация паттернов изображений. Задача автоматической классификации решалась с использованием тестового набора данных «Image Segmentation Data Set» (UCI Re-

pository of Machine Learning), содержащего 20 атрибутов, и представленного в виде таблицы в xlsx формате (рис.3).



region-centroid-col numeric	region-centroid-row	region-pixel-count	short-line-density-5	short-line-density-2	vedge-mean	vegde-sd	hedge-mean	hedge-sd	intensity-mean	rawred-mean	rawblue mean
38	189	9	0	0	1	0.222222	6.22222	33.3185	29.0741	26.3333	35.2222
25	199	9	0	0	1.11111	0.607407	1.05556	0.462963	17.5185	13.1111	17.8889
49	139	9	0	0	0.166667	0.077778	0.333333	0.088889	0.444444	0	1.33333
63	220	9	0	0	3.05556	15.263	3.66667	6.08889	8.18519	6.55556	6.44444

Рис. 3. Web-интерфейс автоматизированной системы анализа данных

Исследователю предоставляется возможность выбора атрибутов для анализа (Feature Selection). После автоматического сопоставления типов данных выбранных атрибутов с шаблонами методов из базы метаданных (рис.4, 5) выбираются релевантные методы анализа [8], которые реализованы, зарегистрированы в сетевой инфраструктуре и к которым для исследователя установлены права доступа.

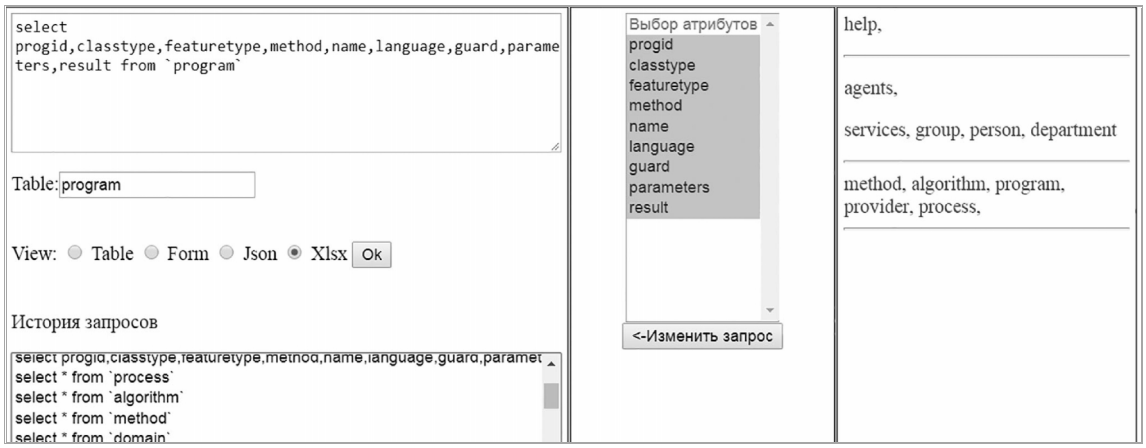


Рис. 4. Web-редактор базы методов и моделей (окно запросов)

Исследователю предоставляется возможность выбора методов из предлагаемого списка и редактирования параметров настроек, заданных по умолчанию (default). На следующем этапе подсистемой планирования и выполнения задач осуществляется применение методов анализа к набору данных в распределенной сетевой среде. Для этого набор данных автоматически преобразуется в стандартный входной формат метода (Attribute-Relation File Format – ARFF, Predictive Model Markup Language – PMML, Comma-Separated Values – CSV). Решение задач исполняется интеллектуальными программными агентами, взаимодействующими с интерфейсами и Web-сервисами. Результаты решения задач анализа данных автоматически сохраняются в базе данных. Для каждого полученного решения вычисляются показатели качества моделирования. Просмотр результатов и выбор лучших моделей по различным критериям осуществляется с использованием Web-интерфейса. Предусмотрена возможность вывода и редактирования отчетов с результатами анализа моделей в Web-редакторе документов. Данные с результатами анализа так же доступны для Web-сервисов в формате JSON.

Сохранить в XLSX

progid	classtype	featuretype	method	language	parameters
1	list	numeric list	j48	java	-C 0.25 -M 2 -t -d -split-percentage
2	list	numeric list	NaiveBayes	java	-t -d -split-percentage
3	list	numeric list	BayesNet	java	-t -d -split-percentage
4	list	numeric list	SMO	java	-C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4" -t -d -split-percentage
5	list	numeric list	SVM	java	-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -t -d -split-percentage
6	list	numeric	NaiveBayesMultinomial	java	-t -d -split-percentage
7	list	numeric	AdaBoostM1	java	-P 100 -S 1 -I 10 -W weka.classifiers.trees.DecisionStump -t -d -split-percentage
8	list	numeric list	LMT	java	-I -1 -M 15 -W 0.0 -t -d -split-percentage
9	list	numeric list	RandomForest	java	-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 -t -d -split-percentage
10	numeric	numeric	LinearRegression	java	-S 0 -R 1.0E-8 -num-decimal-places 4 -t -d -split-percentage

Рис. 5. Web-редактор базы методов и моделей (таблица методов)

На рис.6 изображен фрагмент отчета (в Web-редакторе отчетов) с результатом классификации тестового набора данных, полученным методом Random Forest с точностью 97.3%. Точность указана в графе «test_pc».

segment.xlsx

Источник | Стили | Форматирование | Шрифт | Размер

Результаты тестирования моделей на 50% выборке

date	try	k	featuresc	name	splitpc	aicc	test_pc	kappa	mae	rmse	rae	rrse
2018-04-05 18:31:52	test	20	region-centroid-colnumeric, region-centroid-row, region-pixel-count, short-line-density-5, short-line-density-2, vedge-mean, vegde-sd, hedge-mean, hedge-sd, intensity-mean, rawred-mean, rawblue-mean, rawgreen-mean,	RandomForest	50	-3628.91	97.3333	0.9689	0.0236	0.0867	9.6329	24.7756

Рис. 6. Web-редактор отчетов

Так же был проведен автоматизированный анализ изображений объектов океана с помощью моделей CNN. На рис.7 изображен пример результата классификации паттерна кита с применением TenzorFlow с точностью 97.2%.

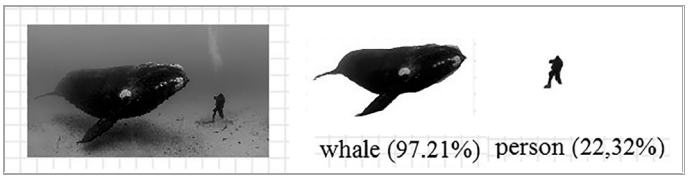


Рис. 7. Классификация паттернов методом конволюционных нейронных сетей (CNN, TenzorFlow)

Задача 2. Прогнозирование успешности обучения студентов. Известные оценки производительности экспертных систем (например, [11]) показывают, что возможно с точностью около 70% прогнозировать эффективность учебной деятельности. На основе численных признаков X, полученных в результате диагностики индивидуальных когнитивных параметров студентов [15] и итоговых баллов академической успеваемости Y различными методами построены модели классификации и регрессионного анализа. Разработаны и исследованы несколько сценариев автоматизированного анализа. Лучший показатель точности прогнозирования был получен в сценарии, основанном на последовательном применении методов кластерного анализа, классификации и регрессионного анализа. Для классификации по результатам диагностики студенты были условно разделены с помощью кластерного анализа априорных данных успешности обучения на 3 условных группы, обозначенных: high (H), average (A) и special (S). Затем для 50%-й обучающей выборки были построены модели классификации. Далее для каждого кластера были сформированы регрессионные модели. Результаты экспериментов с обучающей выборкой (160 студентов) показали точность предсказания 100%, а с тестовой выборкой (60 студентов) 86%. Наилучшие результаты по точности, критерию AIC и критерию интерпретации получены по моделям деревьев решений (J48, LMT). Последовательное применение методов классификации и регрессии на 50%-й тестовой выборке студентов позволяет прогнозировать оценки успешности обучения со среднеквадратическими ошибками (RMSE): 0.15(S),0.06(A), 0.069(H).

Задача 3. Классификация публикаций на примерах тематических областей знаний «Методы обработки изображений и сигналов в гидроакустике», «Инженерия знаний». Проведена предварительная обработка корпусов текстов: графематическая разметка, формирование частотного словаря, анализ частей речи и лемматизация, формирование векторной модели однословных и многословных терминов. Выбраны наиболее значимые по частоте признаки (более 5000). Тексты классифицированы с использованием нескольких моделей. Лучшими по критерию точности оказались деревья решений и байесовские модели. Приоритет отдан байесовским моделям, т. к. модели деревьев решений выделяли значительно меньшее число существенных признаков, что в последствии может привести к снижению точности классификации новых документов.

Заключение

В докладе представлены методы автоматизированного анализа данных для решения задач классификации и регрессионного анализа. Реализация сетевой среды для автоматизированного анализа данных отличается поддержкой стандартов семантического Web, применением технологий распределенных сетевых агентов и сервисов, поддержкой динамического расширения числа методов и масштабированием. Разработка сценариев обработки и анализа данных на логическом уровне моделирования с использованием онтологий создает основу для эффективной реализации в динамической сетевой среде. Из предлагаемых методов анализа данных автоматически выбирается лучший по комплексному критерию, что является особенностью данной системы. Данный подход позволяет сравнивать результаты, полученные разными моделями, учитывая не только точность, но и другие показатели, например, сложность модели (число параметров, признаков) в соответствии с принципом Оккама.

Объединение сценариев с операциями обработки и интеллектуального анализа позволяет эффективно решать разные исследовательски задачи: задачу №1 классификации паттернов изображений в масштабируемой сетевой среде. Разработанный сценарий для 2-й задачи автоматизированного анализа когнитивных параметров обучаемых отличается последовательным применением методов классификации и регрессии, что позволяет повысить точность прогнозирования успешности обучения студентов. Применение многокритериальной оценки качества моделирования и использование критерия «интерпретируемости» человеком (пользователем) позволило выбрать модель деревьев решений из сопоставимых по точности моделей, что позволило реализовать подсистему «объяснения» результатов прогнозирования успешности обучения студентов. Разработка объединенного сценария обработки и анализа текстов для задачи №3 автоматизированной классификации новых документов на принадлежность тематическим областям (на примере «Методов обработки изображений и сигналов в гидроакустике», и «Инженерии знаний») позволяет реализовать автоматизированную подсистему поддержки научных исследований.

Информационные технологии автоматизированного анализа данных реализованы в сетевой программной среде ОнтоМАСТЕР-Ресурс и могут быть применены для проведения научных исследованиях и в процессе обучения студентов.

Работа проводилась при частичной поддержке гранта РНФ, проект № 17-71-20077.

ЛИТЕРАТУРА

1. Guyon I., Chaabane I., et al. A brief review of the ChaLearn AutoML challenge: any-time any-dataset learning without human intervention //Workshop on Automatic Machine Learning. 2016. pp. 21–30.
2. Olson R. S., Bartley N., et al. Evaluation of a tree-based pipeline optimization tool for automating data science //Proceedings of the Genetic and Evolutionary Computation Conference 2016. ACM. 2016. pp. 485–492.
3. Witten I. H., Frank E., et al. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann. 2016. 629 p.
4. Hanif M. H. M., Adewole K. S., et al. Performance Evaluation of Machine Learning Algorithms for Spam Profile Detection on Twitter Using WEKA and RapidMiner //Advanced Science Letters. – 2018. V. 24. No 2. Pp. 1043–1046.
5. Fillbrunn A., Dietz C., et al. KNIME for reproducible cross-domain analysis of life science data //Journal of Biotechnology. 2017. vol. 261. pp. 149–156.
6. Triguero I., González S., et al. KEEL 3.0: an open source software for multi-stage analysis in data mining //International Journal of Computational Intelligence Systems. 2017. vol. 10. no. 1.pp. 1238–1249.
7. Abadi M., Barham P., et al. TensorFlow: A System for Large-Scale Machine Learning//OSDI. 2016. vol. 16. pp. 265–283.
8. Котова Е.Е., Писарев А.С., Писарев И.А. Сетевой метод автоматизации интеллектуального анализа данных в научных исследованиях. Труды XXI Международной конференции по мягким вычислениям и измерениям (SCM'2018). Санкт-Петербург. 2018.
9. Котова Е.Е., Писарев И.А. Метод визуализации терминологических словарей и онтологий. Труды III Международной конференции «Человеческий фактор в сложных технических системах и средах» (Эрго-2018) (Россия, Санкт-Петербург, 4–7 июля 2018 г.). СПб. 2018. С. 240-248.
10. Strecht P., Cruz L., et al. A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance //International Educational Data Mining Society. 2015. pp.392–395.
11. Asif R., Merceron A., et al. Analyzing undergraduate students' performance using educational data mining //Computers & Education. 2017. vol. 113. pp. 177–194.
12. Burnham K. P., Anderson D. R. Multimodel inference: understanding AIC and BIC in model selection//Sociological methods & research. 2004. vol. 33. no. 2. pp. 261–304.
13. Pisarev I. A., Kotova E. E., Pisarev A. S., Stash N. V. Structure of knowledge base of methods for processing hydroacoustic signals //Young Researchers in Electrical and Electronic Engineering (EIConRus), 2018 IEEE Conference of Russian. IEEE, 2018. pp. 1132–1135.
14. Котова Е. Е., Писарев И. А. Представление знаний средствами онтологий в учебном процессе. В сборнике: материалы конференции "Информационные технологии в управлении" (ИТУ-2014) 2014. С. 774–779.
15. Имаев Д.Х., Котова Е.Е. Оценка параметров динамических моделей обучаемых по результатам экспресс-диагностирования. // Известия СПбГЭТУ «ЛЭТИ». Современные технологии образования. 2015. № 1. С. 70–75.

E. E. Kotova, A. S. Pisarev, I. A. Pisarev (Saint Petersburg Electrotechnical University "LETI", St. Petersburg)

Technologies of automated processing and intellectual analysis of unstructured data.

There are presented information technologies for automation of data mining scenarios for application in scientific research and teaching of students. A hybrid database of information resources (documents, images, video, etc.), a metadata base of methods and models, processing and data mining tools in the Web environment have been developed. Metadata describes more than 50 machine learning methods related to the classification of information resources based on the Bayes Theorem, Decision Trees, Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and etc. Examples of scenarios of scientific research are given. Automated data analysis is implemented in OntoMASTER-Resource network software environment on the basis of Semantic Web standards.

МАТЕРИАЛЫ КОНФЕРЕНЦИИ
«ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В УПРАВЛЕНИИ»
(ИТУ-2018)

Подписано в печать 06.11.18. Формат 60×84/8. Бумага офсетная. Гарнитура Таймс и Ариал.
Усл.-печ. л. 73. Тираж 30 экз.

Государственный научный центр Российской Федерации АО «Концерн «ЦНИИ «Электроприбор»
197046, С.-Петербург, ул. Малая Посадская, 30

Отпечатано с готового оригинал-макета, представленного оргкомитетом конференции,
в типографии Издательства Политехнического университета.
195251, С.-Петербург, Политехническая ул., 29