

Software & Systems
Programmnye produkty i sistemy
2017, vol. 30, no. 1

Print ISSN 0236-235X

Online ISSN 2311-2735

Программные продукты и системы

DOI 10.15827/0236-235X

/ О структурированном
хранении данных

/ Методы RL-обучения
на основе
темпоральных
различий

/ Схема
многопроцессорной
обработки
3-D данных

/ Методы
автоматической
классификации
текстов



№
1
ТОМ 30
2017

Научно-исследовательский институт
«Центрпрограммсистем»

Программные продукты и системы

МЕЖДУНАРОДНЫЙ НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ

2017, том 30, № 1
(год издания тридцатый)

Главный редактор
С.В. ЕМЕЛЬЯНОВ, академик РАН

Тверь

SOFTWARE & SYSTEMS

(PROGRAMMNYE PRODUKTY I SISTEMY)

International research and practice journal

2017, vol. 30, no. 1

Editor-in-Chief
S.V. EMEL'YANOV, Academician of the Russian Academy of Sciences

Tver
Russian Federation

Research Institute CENTERPROGRAMSYSTEM

© ПРОГРАММНЫЕ ПРОДУКТЫ И СИСТЕМЫ

Международный научно-практический журнал

2017. Т. 30. № 1

DOI: 10.15827/0236-235X

Главный редактор

С.В. ЕМЕЛЬЯНОВ,
академик РАН (г. Москва, Россия)

Научные редакторы:

Н.А. СЕМЕНОВ, д.т.н., профессор ТвГТУ
(г. Тверь, Россия)

А.П. ЕРЕМЕЕВ, д.т.н., профессор МЭИ
(г. Москва, Россия)

Издатель НИИ «Центрпрограммсистем»
(г. Тверь, Россия)

Учредители: МНИИПУ (г. Москва, Россия),
Главная редакция международного журнала
«Проблемы теории и практики управления» (г. Москва, Россия),
Закрытое акционерное общество
«Научно-исследовательский институт
«Центрпрограммсистем» (г. Тверь, Россия)

Журнал зарегистрирован в Комитете Российской Федерации
по печати 26 июня 1995 г.

Регистрационное свидетельство № 013831

Подписной индекс в каталоге
Агентства «Роспечать» **70799**

ISSN 0236-235X (печатн.)

ISSN 2311-2735 (онлайн)

МЕЖДУНАРОДНАЯ РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Семенов Н.А. – д.т.н., профессор Тверского государственного технического университета, заместитель главного редактора (г. Тверь, Россия)

Решетников В.Н. – д.ф.-м.н., профессор Московского авиационного института (национального исследовательского университета), заместитель главного редактора (г. Москва, Россия)

Арефьев И.Б. – д.т.н., профессор Морской академии Польши (г. Щецин, Польша)

Афанасьев А.П. – д.ф.-м.н., профессор Московского физико-технического института (технического университета), заведующий Центром распределенных вычислений Института проблем передачи информации РАН (г. Москва, Россия)

Баламетов А.Б. – д.т.н., профессор Азербайджанского научно-исследовательского и проектно-изыскательского института энергетики (г. Баку, Азербайджан)

Батыршин И.З. – д.т.н., профессор Мексиканского института нефти (г. Мехико, Мексика)

Вагин В.Н. – д.т.н., профессор Московского энергетического института (технического университета)
(г. Москва, Россия)

Голенков В.В. – д.т.н., профессор Белорусского государственного университета информатики и радиоэлектроники
(г. Минск, Беларусь)

Еремеев А.П. – д.т.н., профессор Московского энергетического института (технического университета)
(г. Москва, Россия)

Котов А.С. – кандидат наук, ассистент профессора университета Уэйна (штат Мичиган) (г. Детройт, США)

Кузнецов О.П. – д.т.н., профессор Института проблем управления РАН (г. Москва, Россия)

Курейчик В.М. – д.т.н., профессор Инженерно-технологической академии Южного федерального университета
(г. Таганрог, Россия)

Лисецкий Ю.М. – к.т.н., генеральный директор «S&T Ukraine» (г. Киев, Украина)

Мамросенко К.А. – к.т.н., доцент Московского авиационного института (национального исследовательского университета), руководитель Центра визуализации и спутниковых информационных технологий НИИСИ РАН (г. Москва, Россия)

Мейер Б. – доктор наук, профессор, заведующий кафедрой Высшей политехнической школы – ЕТН (г. Цюрих, Швейцария)

Науен Тхань Нгу – д.ф.-м.н., профессор, проректор Ханойского открытого университета (г. Ханой, Вьетнам)

Николов Р.В. – доктор наук, профессор Университета библиотековедения и информационных технологий Софии
(г. София, Болгария)

Осипов Г.С. – д.ф.-м.н., профессор, заместитель директора Института системного анализа РАН (г. Москва, Россия)

Палюх Б.В. – д.т.н., профессор Тверского государственного технического университета (г. Тверь, Россия)

Рахманов А.А. – д.т.н., профессор, заместитель генерального директора Концерна «РТИ Системы» (г. Москва, Россия)

Серов В.С. – д.ф.-м.н., профессор Университета прикладных наук Оулу (г. Оулу, Финляндия)

Сотников А.Н. – д.ф.-м.н., профессор, Межведомственный суперкомпьютерный центр РАН (г. Москва, Россия)

Сулейманов Д.Ш. – академик АН Республики Татарстан, д.т.н., профессор Казанского государственного технического университета (г. Казань, Республика Татарстан, Россия)

Тарасов В.Б. – к.т.н., доцент Московского государственного технического университета им. Н.Э. Баумана (г. Москва, Россия)

Таратухин В.В. – доктор философии, управляющий директор Европейского исследовательского центра в области информационных систем (ERCIS) Вестфальского университета им. Вильгельма (г. Мюнстер, Германия)

Хорошевский В.Ф. – д.т.н., профессор Московского физико-технического института (технического университета)
(г. Москва, Россия)

Язенин А.В. – д.ф.-м.н., профессор Тверского государственного университета (г. Тверь, Россия)

АССОЦИИРОВАННЫЕ ЧЛЕНЫ РЕДАКЦИИ

Московский энергетический институт (технический университет), г. Москва, Россия

Технологический институт Южного федерального университета, г. Таганрог, Россия

Тверской государственный технический университет, г. Тверь, Россия

Научно-исследовательский институт «Центрпрограммсистем», г. Тверь, Россия

АДРЕС РЕДАКЦИИ

Россия, 170024, г. Тверь, пр. 50 лет Октября, 3а

Телефон (482-2) 39-91-49

Факс (482-2) 39-91-00

E-mail: red@cps.tver.ru

www.swsys.ru

Подписано в печать 28.02.2017 г.

Отпечатано ООО ИПП «Фактор и К»

Россия, 170028, г. Тверь, ул. Лукина, д. 4, стр. 1

Выпускается один раз в квартал. Общее количество выпусков 117.

Год издания тридцатый. Формат 60×84 1/8. Объем 160 стр.

Заказ № 4. Тираж 1000 экз. Цена 257,40 руб.

Автор статьи отвечает за подбор, оригинальность и точность приводимого фактического материала.
Авторские гонорары не выплачиваются. При перепечатке материалов ссылка на журнал обязательна.

Editor-in-chief

S.V. Emelyanov, Academician of the Russian Academy of Sciences
(Moscow, Russian Federation)

Science editors:

N.A. Semenov, Dr.Sc. (Engineering), Professor TvSTU
(Tver, Russian Federation)

A.P. Ereemeev, Dr.Sc. (Engineering), Professor MPEI
(Moscow, Russian Federation)

INTERNATIONAL EDITORIAL BOARD

Semenov N.A. – Dr.Sc. (Engineering), Professor of Tver State Technical University, Deputy Editor-in-Chief
(Tver, Russian Federation)

Reshetnikov V.N. – Dr.Sc. (Physics and Mathematics), Professor of Moscow Aviation Institute (National Research University),
Deputy Editor-in-Chief (Moscow, Russian Federation)

Arfeev I.B. – Dr.Sc. (Engineering), Professor of Poland Szczecin Maritime Academy (Szczecin, Poland)

Afanasiev A.P. – Dr.Sc. (Physics and Mathematics), Professor of Moscow Institute of Physics and Technology,
Head of Centre for Distributed Computing of Institute for Information Transmission Problems (Moscow, Russian Federation)

Balametov A.B. – Azerbaijan Scientific-Research & Design-Prospecting Power Engineering Institute (Baku, Azerbaijan)

Batyrshin I.Z. – Dr.Sc. (Engineering), Professor of Mexican Petroleum Institute (Mexico City, Mexico)

Vagin V.N. – Dr.Sc. (Engineering), Professor of Moscow Power Engineering Institute (Technical University)
(Moscow, Russian Federation)

Golenkov V.V. – Dr.Sc. (Engineering), Professor of Belarusian State University of Informatics and Radioelectronics
(Minsk, Republic of Belarus)

Ereemeev A.P. – Dr.Sc. (Engineering), Professor of Moscow Power Engineering Institute (Technical University)
(Moscow, Russian Federation)

Kotov A.S. – Ph.D. (Computer Science), Assistant Professor, Wayne State University (Detroit, MI, USA)

Kuznetsov O.P. – Dr.Sc. (Engineering), Professor of the Institute of Control Sciences of the Russian Academy of Sciences
(Moscow, Russian Federation)

Kureichik V.M. – Dr.Sc. (Engineering), Professor of Academy of Engineering and Technology Southern Federal University
(Taganrog, Russian Federation)

Lisetskiy Yu.M. – Ph.D. Tech.Sc., CEO of "S&T Ukraine" (Kiev, Ukraine)

Mamrosenko K.A. – Ph.D. (Engineering), Associate Professor of Moscow Aviation Institute (National Research University),
Head of Center of Visualization and Satellite Information Technologies SRISA RAS (Moscow, Russian Federation)

Meyer B. – Dr.Sc., Professor, Head of Department in Swiss Federal Institute of Technology in Zurich, ETH
(Zurich, Switzerland)

Nguyen Thanh Nghi – Dr.Sc. (Physics and Mathematics), Professor, Vice-Principal of Hanoi Open University (Hanoi, Vietnam)

Nikolov R.V. – Full Professor of the University of Library Studies and Information Technology (Sofia, Bulgaria)

Osipov G.S. – Dr.Sc. (Physics and Mathematics), Professor, Deputy of the Principal of Institute of Systems Analysis
of the Russian Academy of Sciences (Moscow, Russian Federation)

Palyukh B.V. – Dr.Sc. (Engineering), Professor of Tver State Technical University (Tver, Russian Federation)

Rakhmanov A.A. – Dr.Sc. (Engineering), Professor, Deputy of the CEO of Concern "RTI Systems"
(Moscow, Russian Federation)

Serov V.S. – Dr.Sc. (Physics and Mathematics), Professor of the Oulu University of Applied Sciences (Oulu, Finland)

Sotnikov A.N. – Dr.Sc. (Physics and Mathematics), Professor, Joint Supercomputer Center of the Russian Academy
of Sciences (Moscow, Russian Federation)

Suleimanov D.Sh. – Academician of TAS, Dr.Sc. (Engineering), Professor of Kazan State Technical University
(Kazan, Republic of Tatarstan, Russian Federation)

Tarassov V.B. – Ph.D. (Engineering), Associate Professor of Bauman Moscow State Technical University
(Moscow, Russian Federation)

Taratoukhine V.V. – Ph.D. (Engineering), Dr.Ph., Managing Director of the Competence Centre ERP and ERCIS Lab
Russia of the ERCIS (Muenster, Germany)

Khoroshevsky V.F. – Dr.Sc. (Engineering), Professor of Moscow Institute of Physics and Technology
(Moscow, Russian Federation)

Yazenin A.V. – Dr.Sc. (Physics and Mathematics), Professor of Tver State University (Tver, Russian Federation)

ASSOCIATED EDITORIAL BOARD MEMBERS

Moscow Power Engineering Institute (Technical University), Moscow, Russian Federation

Technology Institute at Southern Federal University, Taganrog, Russian Federation

Tver State Technical University, Tver, Russian Federation

Research Institute CENTERPROGRAMSYSTEM, Tver, Russian Federation

EDITORIAL OFFICE ADDRESS

50 let Oktyabrya Ave. 3a, Tver, 170024, Russian Federation
Phone: (482-2) 39-91-49 **Fax:** (482-2) 39-91-00
E-mail: red@cps.tver.ru
www.swsys.ru

Passed for printing 28.02.2017
Printed in printing-office "Faktor i K"
Lukina St. 4/1, Tver, 170028, Russian Federation
Published quarterly. 30th year of publication
Format 60×84 1/8. Circulation 1000 copies
Prod. order № 4. Wordage 160 pages. Price 257,40 rub.

Вниманию авторов

Международный журнал «Программные продукты и системы» публикует материалы научного и научно-практического характера по новым информационным технологиям, результаты академических и отраслевых исследований в области использования средств вычислительной техники. Практикуются выпуски тематических номеров по искусственному интеллекту, системам автоматизированного проектирования, по технологиям разработки программных средств и системам защиты, а также специализированные выпуски, посвященные научным исследованиям и разработкам отдельных вузов, НИИ, научных организаций.

Решением Президиума Высшей аттестационной комиссии (ВАК) Министерства образования и науки РФ международный журнал «Программные продукты и системы» внесен в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней кандидата и доктора наук.

Информация об опубликованных статьях по установленной форме регулярно предоставляется в систему Российского индекса научного цитирования (РИНЦ), в CrossRef и в другие базы и электронные библиотеки.

Условия публикации

К рассмотрению принимаются ранее нигде не опубликованные материалы, соответствующие тематике журнала (специализация 05.13.XX – Информатика, вычислительная техника и управление) и отвечающие редакционным требованиям.

Работа представляется в электронном виде в формате Word. При обилии сложных формул обязательно наличие статьи и в формате PDF. Формулы должны быть набраны в редакторе формул Word (Microsoft Equation или MathType). Объем статьи вместе с иллюстрациями – не менее 10 000 знаков. Диаграммы, схемы, графики должны быть доступными для редактирования (Word, Visio, Excel). Все иллюстрации для полиграфического воспроизведения представляются в черно-белом варианте. Цветные, тонированные, отсканированные, не подлежащие редактированию средствами Word рисунки и экранные формы следует присылать в хорошем качестве для их дополнительного размещения на сайте журнала в макете статьи с доступом по ссылке. (Публикация материалов с использованием гипертекста, графики, аудио-, видео-, программных средств и др. возможна в электронном издании «Программные продукты, системы и алгоритмы», сайт www.swsys-web.ru.) Заголовок должен быть информативным; сокращения, а также терминологию узкой тематики желательно в нем не использовать. Количество авторов на одну статью – не более 4, количество статей одного автора в номере, включая соавторство, – не более 2. Список литературы, наличие которого обязательно, должен включать не менее 10 пунктов.

Необходимы также содержательная структурированная аннотация (не менее 200 слов), ключевые слова (7–10) и индекс УДК. Название статьи, аннотация и ключевые слова должны быть переведены на английский язык (машинный перевод недопустим), а фамилии авторов, названия и юридические адреса организаций (если нет официального перевода), пристатейные списки литературы – транслитерированы по стандарту BGN/PCGN.

Вместе со статьей следует прислать сопроводительное письмо-рекомендацию в произвольной форме, экспертное заключение, лицензионное соглашение, а также сведения об авторах: фамилия, имя, отчество, название и юридический адрес организации, должность, ученые степень и звание (если есть), контактный телефон, электронный адрес, почтовый адрес для отправки бесплатного авторского экземпляра журнала.

Порядок рецензирования

Все статьи, поступающие в редакцию (соответствующие тематике и оформленные согласно требованиям к публикации), подлежат обязательному рецензированию в течение месяца с момента поступления.

В редакции есть устоявшийся коллектив рецензентов, среди которых члены международной редколлегии журнала, эксперты из числа крупных специалистов в области информатики и вычислительной техники ведущих вузов страны, а также ученые и специалисты НИИ «Центрпрограммсистем» (г. Тверь).

Рецензирование проводится конфиденциально. Автору статьи предоставляется возможность ознакомиться с текстом рецензии. При необходимости статья отправляется на доработку.

Рецензии обсуждаются на заседаниях рабочей группы, состоящей из членов научного совета журнала. Заседания проводятся раз в месяц в НИИ «Центрпрограммсистем» (г. Тверь), где принимается решение о целесообразности публикации статьи.

Статьи, одобренные редакционным советом, публикуются бесплатно в течение года с момента одобрения, а отправленные на доработку – с момента поступления после устранения замечаний.

Редакция международного журнала «Программные продукты и системы» в своей работе руководствуется сводом правил Кодекса этики научных публикаций, разработанным и утвержденным Комитетом по этике научных публикаций.

УДК 004.5; 004.8

DOI: 10.15827/0236-235X.117.005-011

Дата подачи статьи: 22.07.16

2017. Т. 30. № 1. С. 5–11

ПРОЕКТИРОВАНИЕ ИНТЕРФЕЙСА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ С ИСПОЛЬЗОВАНИЕМ ЭЛЕМЕНТОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Т.М. Зубкова, д.т.н., профессор, bars87@mail.ru

(Оренбургский государственный университет, просп. Победы, 13, г. Оренбург, 460018, Россия);

Е.Н. Наточая, к.п.н., доцент, en_ischa@mail.ru

*(Российская академия народного хозяйства и государственной службы при Президенте
Российской Федерации (Оренбургский филиал), ул. Курача, 26, г. Оренбург, 460000, Россия)*

Для разработки качественного ПО необходимо в техническом задании отразить все требования и пожелания заказчика, чтобы у него и у исполнителя сложилось единое представление о будущем программном продукте. Одним из вариантов достижения данного взаимопонимания является разработка прототипа пользовательского интерфейса.

В статье описана методика подбора альтернативного варианта шаблона интерфейса, использующая такие методы искусственного интеллекта, как экспертная оценка и теория нечетких множеств. На основе индивидуальных характеристик пользователей можно разделить на пять групп: новичок, обычный, уверенный, квалифицированный, администратор. Выявлены основные параметры индивидуальных характеристик, по которым можно классифицировать пользователей при проектировании интерфейсов: компьютерная грамотность, системный опыт, опыт работы с подобными программами, машинопись, мышление, память, моторика, дальтонизм, концентрация внимания, эмоциональная устойчивость.

В статье описано программное и математическое обеспечение для решения задач интеллектуального проектирования пользовательского интерфейса. Поставленная задача выполняется в три этапа. На первом этапе – «Формирование и оценка компетентности группы экспертов» – определяются характеристики экспертов. Количественное описание характеристик экспертов основывается на вычислении относительных коэффициентов компетентности по результатам высказываний специалистов о составе экспертной группы. На втором этапе – «Групповая экспертная оценка объектов при непосредственном оценивании» – определяются рекуррентные отношения для итераций. Третий этап – «Построение нечеткой модели на бинарных нечетких отношениях» – оперирует двумя нечеткими множествами: совокупность групп пользователей и множество шаблонов интерфейсов, максимально эффективных для пользователей с данными характеристиками. Входными данными нечеткой модели являются выделенные нечеткие множества, а выходными – степени соответствия шаблонов интерфейса пользователям.

На основе предложенной методики автоматизирован процесс проектирования пользовательского интерфейса с целью повышения объективности и оперативности решений, принимаемых разработчиками ПО.

Ключевые слова: ПО, пользовательский интерфейс, шаблон интерфейса, прототип интерфейса, техническое задание, экспертная оценка, база знаний, нечеткие отношения.

При проектировании программных продуктов необходимо учитывать требования, пожелания, а также знания и возможности потенциального контингента пользователей, которые отражаются в техническом задании на ПО. Для получения единого представления о создаваемом ПО предусматривается активное взаимодействие разработчика с заказчиком. Заказчики разрабатывают концепцию (часто подсознательную и неполную) того, как их приложение будет работать. Разработчикам же необходимо учитывать аппаратные ресурсы, базовое ПО, операционную систему и другое. Кроме того, требуется знать возможности целевой аудитории пользователей, варианты использования ПО или сценарии работы, которые формируются из пользовательских историй. Одним из способов взаимодействия с заказчиком и получения единого видения ПО является разработка прототипа интерфейса [1]. Прототип пользовательского интерфейса представляет собой макет (черновую, пробную версию) интерфейса, разрабатываемый с целью проверки пригодности предлагаемых для применения концепций, технологических решений, а также для представления программы заказчику на ранних стадиях процесса разработки.

Постановка задачи

Задача заключается в разработке программного средства с функцией поддержки принятия решения по подборке альтернативного варианта шаблона интерфейса на основании экспертной оценки и теории нечетких множеств. Для автоматизации принятия решения необходимо разработать шаблоны пользовательского интерфейса с учетом целевой аудитории [2].

Под шаблоном пользовательского интерфейса будем понимать сочетание одного или нескольких элементов, широко используемых для предоставления навигации, команд и содержимого, которые используются большинством приложений и служат основой для разработки собственного пользовательского интерфейса.

Характеристиками контингента пользователей являются их уровень знаний, физические и психологические качества. Каждая из характеристик оценивается по трехбалльной шкале: высокий, низкий и средний уровень развития.

С учетом разных градаций выделенных характеристик пользователи разделены на пять групп (табл. 1).

Таблица 1
Характеристики групп пользователей
Table 1
User group dimensions

Характеристика	Группа пользователей				
	Новичок	Обычный	Уверенный	Квалифицированный	Программист, хакер, администратор
Компьютерная грамотность	0,1	0,4	0,7	0,8	1
Системный опыт	0,1	0,5	0,7	0,8	1
Опыт работы с подобными программами	0,1	0,5	0,6	0,7	0,9
Машинопись	0,7	0,7	0,8	0,9	0,9
Мышление	0,3	0,6	0,7	0,8	0,9
Память	0,2	0,5	0,6	0,8	1
Моторика	0,6	0,7	0,8	0,9	0,9
Дальтонизм	0,1	0,1	0,1	0,1	0,1
Концентрация внимания	0,3	0,7	0,8	0,9	0,1
Эмоциональная устойчивость	0,2	0,5	0,6	0,8	0,9

Декомпозиция функций программной системы представлена в нотации IDEF0 (Integrated DEFinition) на рисунке 1.

Входными данными программной системы являются характеристики групп пользователей, требования заказчика, лицензия разработчика, лицензия заказчика, шаблоны интерфейсов.

В качестве управляющего воздействия выступают Федеральный закон «О персональных данных», ГОСТ19.201-78, экспертные оценки специалистов в данной предметной области, нечеткие отношения.

На выходе программной системы формируется фрагмент технического задания на разработку ПО с прототипом интерфейса.

Реализация поставленной задачи

Входные данные для задачи проектирования интерфейса программных средств характеризуются той или иной степенью неопределенности, обусловленной неполнотой, внутренней противоречивостью, неоднозначностью, и представляют собой приближенные количественные или качественные оценки параметров процессов проектирования и управления проектированием [3].

Таким образом, так как исходные данные задачи трудно формализуемы, целесообразно применить один из методов искусственного интеллекта, основанный на нечеткой логике.

Нечеткие алгоритмы, оперирующие лингвистическими переменными, значения которых задаются нечеткими множествами, удобны для описания слабо формализуемых процессов. Такие алгоритмы

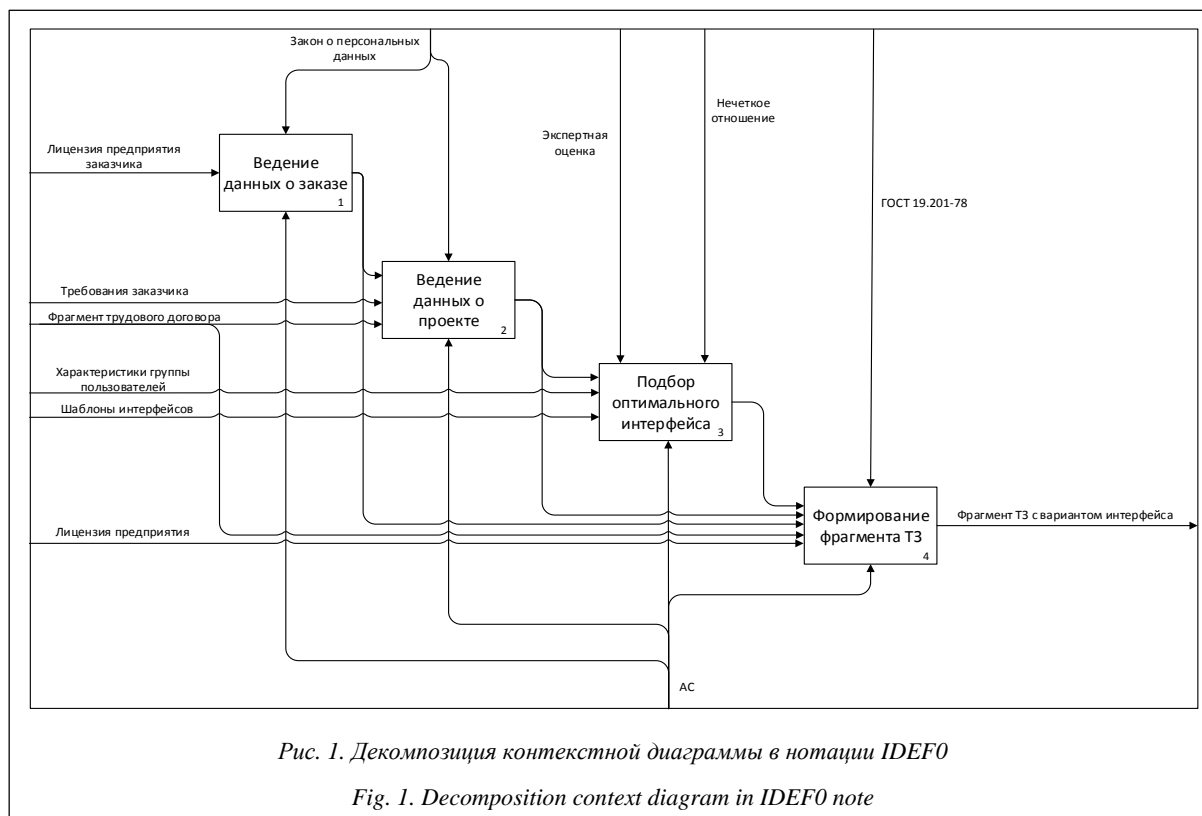


Рис. 1. Декомпозиция контекстной диаграммы в нотации IDEF0

Fig. 1. Decomposition context diagram in IDEF0 note

интуитивно более понятны. Их автоматизация позволяет повысить объективность и оперативность решений, принимаемых разработчиком ПО [4].

Методология нечеткого логического вывода достаточно успешно применяется при построении систем управления объектами [5], в частности, при разработке компонентов ПО [6, 7] и минимизации рисков программных проектов [8, 9].

Этап 1. Формирование и оценка компетентности группы экспертов.

Экспертная оценка – процедура получения оценки проблемы на основе мнения специалистов с целью принятия решения. Экспертное оценивание предполагает создание некоего разума, обладающего большими способностями по сравнению с отдельным человеком. Основной сложностью этого метода является подбор экспертов, которые должны иметь опыт в соответствующих решаемым задачам областях. При подборе экспертов следует учитывать личную заинтересованность, которая может стать существенным препятствием для получения объективного суждения.

При формировании группы экспертов на стадии выявления знаний учитывались следующие характеристики:

- компетентность (степень квалификации эксперта в данной области знаний);
- креативность (способность решать творческие задачи);
- отношение к экспертизе (негативное или пассивное отношение, занятость, существенно влияющие на качество работы эксперта в группе);
- конформизм (подверженность влиянию авторитетов, при котором их мнение может подавлять лиц, обладающих более высокой компетентностью);
- коллективизм и самокритичность.

Авторами реализован один из возможных путей количественного описания характеристик эксперта, основанный на вычислении относительных коэффициентов компетентности по результатам высказываний специалистов о составе экспертной группы.

Суть методики сводится к тому, что ряду специалистов предлагается высказать мнение о списочном составе экспертной группы. Если в этом списке появляются лица, не вошедшие в исходный список, им тоже предлагается назвать специалистов для участия в экспертизе. После нескольких этапов будет получен достаточно полный список кандидатов в группу [10].

По результатам опроса составляется матрица, по строкам и столбцам которой записываются фамилии экспертов, а элементами таблицы являются переменные:

$$x_{ij} = \begin{cases} 1, & \text{если } j\text{-й эксперт назвал } i\text{-го;} \\ 0, & \text{если } j\text{-й эксперт не назвал } i\text{-го.} \end{cases}$$

При этом эксперт может включать или не включать себя в экспертную группу (то есть $x_{ij} = 0$ или $x_{ij} = 1$).

По данной таблице можно вычислить относительные коэффициенты компетентности, используя алгоритм решения задач о лидере.

Относительные коэффициенты компетентности h -порядка для каждого эксперта имеют следующий вид:

$$K_i^h = \frac{\sum_{j=1}^m x_{ij} k_j^{h-1}}{\sum_{i=1}^m \sum_{j=1}^m x_{ij} k_j^{h-1}}, i = \overline{1, m}, h = 1, 2, \dots,$$

где m – число экспертов в списке (размерность матрицы $\|x_{ij}\|$); h – номер порядка коэффициента компетентности.

Коэффициенты компетентности нормированы так, что их сумма равна единице:

$$\sum_{i=1}^m k_i^h = 1, h = 1, 2, \dots \quad (1)$$

По формуле (1) можно вычислить значение компетентности для различных порядков, начиная с первого.

При $h = 1$ выражение (1) будет иметь вид

$$K_i^1 = \frac{\sum_{j=1}^m x_{ij}}{\sum_{i=1}^m \sum_{j=1}^m x_{ij}}, i = \overline{1, m}. \quad (2)$$

Таким образом, коэффициент компетентности первого порядка – это относительное число экспертов, высказавшихся за включение i -го эксперта в группу.

Относительный коэффициент компетентности второго порядка получаем из (1) для $h = 2$ при условии, что $k_j^1, j = \overline{1, m}$, определены по (2):

$$K_i^2 = \frac{\sum_{j=1}^m x_{ij} k_j^1}{\sum_{i=1}^m \sum_{j=1}^m x_{ij} k_j^1}, i = \overline{1, m}.$$

Коэффициенты второго порядка представляют собой относительное количество голосов взвешенных коэффициентов компетентности первого порядка.

Последовательно вычисляя относительные коэффициенты компетентности более высокого порядка, можно убедиться, что процесс быстро сходится после 3-4 вычислений, то есть относительные коэффициенты быстро стабилизируются [10].

В общем случае коэффициенты относительной компетентности определяются как

$$k_i = \lim_{h \rightarrow \infty} k_i^h, \sum_{i=1}^n k_i = 1.$$

Этап 2. Групповая экспертная оценка объектов при непосредственном оценивании.

Пусть m экспертов провели оценку n объектов по l показателям. Результаты оценивания представлены величинами x_{ij}^h , где i – номер объекта; j – номер эксперта; h – номер показателя. Величины x_{ij}^h , полученные методом непосредственного оценивания, представляют собой числа из некоторого отрезка числовой оси или баллы.

В качестве групповой оценки для каждого из объектов можно принять среднее взвешенное значение его оценки:

$$x_i = \sum_{h=1}^l \sum_{j=1}^m q_h x_{ij}^h k_j, i = 1, 2, \dots, n,$$

где q_h – коэффициенты весов показателей сравнения объектов; k_j – коэффициенты компетентности экспертов.

Величины q_h и k_j являются нормированными, то есть $\sum_{h=1}^l q_h = 1, \sum_{j=1}^m k_j = 1$.

Коэффициенты q_h могут быть определены экспертным путем как средний коэффициент веса h -го показателя по всем экспертам, то есть $q_h = \sum_{j=1}^m q_{hj} k_j$.

При $h=1$ алгоритм вычисления групповых оценок и коэффициентов компетентности экспертов имеет вид:

а) начальные условия при $t=0$:

$$k_j^0 = \frac{1}{m}, j = 1, m, \text{ то есть начальное значение ко-}$$

эффициентов компетентности для всех экспертов принимается одинаковым;

б) рекуррентные соотношения для $t=1, 2, 3, \dots$ представлены в таблице 2;

Таблица 2
Рекуррентные отношения для итераций
Table 2
Recurrence formula for iterations

Аналитическое описание	Расчетная величина
$x_i^t = \sum_{j=1}^m x_{ij}^t k_j^{t-1},$ $i = 1, n$	Групповая оценка для i -го объекта на t -м шаге на основе индивидуальных оценок x_{ij}
$\lambda^t = \sum_{i=1}^n \sum_{j=1}^m x_{ij}^t x_{ij}$	Нормировочный коэффициент
$k_j^t = \frac{1}{\lambda^t} \sum_{i=1}^n x_{ij}^t x_{ij},$ $j = 1, m-1$	Коэффициенты компетентности j -го эксперта на t -м шаге
$k_m^t = 1 - \sum_{j=1}^{m-1} k_j^t$	Коэффициенты компетентности m -го эксперта из условия нормировки [10]

в) признак окончания итерационного процесса:
 $\max \left(\left| x_i^t - x_i^{t-1} \right| \right) < E.$

Этап 3. Построение нечеткой модели на бинарных нечетких отношениях.

Пусть имеются два множества: совокупность групп пользователей и совокупность шаблонов интерфейсов, которые нужно максимально эффективно подобрать для пользователей с заданными характеристиками. Таким образом, входными данными являются указанные множества, а выходными – степени соответствия шаблонов интерфейсов пользователям.

В общем случае нечетким отношением или, точнее, нечетким k -арным отношением, заданным на множествах (универсумах) X_1, X_2, \dots, X_k , называется некоторое фиксированное нечеткое подмножество декартова произведения этих универсумов. Другими словами, если обозначить произвольное нечеткое отношение через Q , то по определению $Q = \{ \langle x_1, x_2, \dots, x_k \rangle, \mu_Q(\langle x_1, x_2, \dots, x_k \rangle) \}$, где $\mu_Q(\langle x_1, x_2, \dots, x_k \rangle)$ – функция принадлежности данного нечеткого отношения, которая определяется как отображение $\mu_Q: X_1 \times X_2 \times \dots \times X_k \rightarrow [0, 1]$. Здесь через $\langle x_1, x_2, \dots, x_k \rangle$ обозначен кортеж из k элементов, каждый из которых выбирается из своего универсума: $x_1 \in X_1, x_2 \in X_2, \dots, x_k \in X_k$.

Бинарное нечеткое отношение задается на базисных множествах X_1, X_2 и определяется как нечеткое отношение $Q = \{ \langle x_i, x_j \rangle, \mu_Q(\langle x_i, x_j \rangle) \}$. Здесь $\mu_Q(\langle x_i, x_j \rangle)$ – функция принадлежности бинарного нечеткого отношения, которая определяется как отображение $\mu_Q: X_1 \times X_2 \rightarrow [0, 1]$, а через $\langle x_i, x_j \rangle$ обозначен кортеж из двух элементов, при этом $x_1 \in X_1, x_2 \in X_2$.

Кроме того, для построения решения задачи необходимы композиции нечетких бинарных отношений.

Пусть Q и R – конечные или бесконечные бинарные нечеткие отношения. Причем нечеткое отношение $Q = \{ \langle x_i, x_j \rangle, \mu_Q(\langle x_i, x_j \rangle) \}$ задано на декартовом произведении универсумов $X_1 \times X_2$, а нечеткое отношение $R = \{ \langle x_j, x_k \rangle, \mu_R(\langle x_j, x_k \rangle) \}$ – на декартовом произведении универсумов $X_2 \times X_3$. Нечеткое бинарное отношение, заданное на декартовом произведении $X_1 \times X_3$ и обозначаемое через $Q \circ R$, называется композицией бинарных нечетких отношений Q и R , а его функция принадлежности определяется следующим выражением:

$$\mu_{Q \circ R}(\langle x_i, x_k \rangle) = \max \left\{ \min \left\{ \mu_Q(\langle x_i, x_j \rangle), \mu_R(\langle x_j, x_k \rangle) \right\} \right\},$$

$$(\forall \langle x_i, x_k \rangle \in X_1 \times X_3).$$

Определенную таким образом композицию бинарных нечетких отношений называют (max-min)-композицией или максиминной сверткой нечетких отношений.

Нечеткое бинарное отношение, заданное на декартовом произведении $X_1 \times X_3$ и обозначаемое че-

рез $Q \circ R$, называется (max-*)-композицией бинарных нечетких отношений Q и R , если его функция принадлежности определяется следующим выражением:

$$\mu_{Q \circ R}(\langle x_i, x_k \rangle) = \max_{x_j \in X_3} \left\{ \mu_Q(\langle x_i, x_j \rangle) * \mu_R(\langle x_j, x_k \rangle) \right\},$$

$$(\forall \langle x_i, x_k \rangle \in X_1 \times X_3).$$

В частности, если в этом выражении вместо операции «*» использовать операцию алгебраического умножения, получим определение (max-prod)-композиции.

Нечеткое бинарное отношение, заданное на декартовом произведении $X_1 \times X_3$ и обозначаемое через $Q \circ R$, называется (min-max)-композицией бинарных нечетких отношений Q и R , если его функция принадлежности определяется следующим выражением:

$$\mu_{Q \circ R}(\langle x_i, x_k \rangle) = \min_{x_j \in X_3} \left\{ \max \left\{ \mu_Q(\langle x_i, x_j \rangle), \mu_R(\langle x_j, x_k \rangle) \right\} \right\},$$

$$(\forall \langle x_i, x_k \rangle \in X_1 \times X_3).$$

С учетом введенных понятий построим нечеткую модель, основанную на двух бинарных нечетких отношениях S и T . Первое из этих нечетких отношений строится на двух базисных множествах X и Y , а второе – на двух базисных множествах Y и Z . Здесь $X = \{x_1, x_2, x_3, x_4, x_5\}$ описывает множество интерфейсов, $Y = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}\}$ – множество характеристик пользователей, а $Z = \{z_1, z_2, z_3, z_4, z_5\}$ – множество групп пользователей.

В данном контексте нечеткое отношение S содержательно описывает характеристики интерфейсов, а T – характеристики пользователей.

Элементы универсумов имеют следующий содержательный смысл:

1) x_1 – шаблон интерфейса 1, x_2 – шаблон интерфейса 2, x_3 – шаблон интерфейса 3, x_4 – шаблон интерфейса 4, x_5 – шаблон интерфейса 5;

2) y_1 – компьютерная грамотность, y_2 – системный опыт, y_3 – опыт работы с подобными программами, y_4 – машинопись, y_5 – мышление, y_6 – память, y_7 – моторика, y_8 – дальтонизм, y_9 – концентрация внимания, y_{10} – эмоциональная устойчивость;

3) z_1 – новичок, z_2 – обычный пользователь, z_3 – уверенный пользователь, z_4 – квалифицированный, z_5 – программист, хакер, администратор.

Для определения соответствия интерфейса группе пользователей воспользовались композициями исходных нечетких отношений. Так, (max-min)- и (max-prod)-композиции дают информацию о степени соответствия шаблона интерфейса группе пользователей, а (min-max)-композиция позволяет определить шаблон, который не подходит для данной группы пользователей [11].

Таким образом, применяются три модели: max-min, max-prod, min-max.

Способы для определения результата композиции нечетких отношений могут быть следующими.

Max-min-композиция, или максимальная нечеткая свертка: $\mu_B(y) = \max_{x \in X} \left\{ \min \left\{ \mu_A(x), \mu_Q(\langle x, y \rangle) \right\} \right\}$.

Max-prod-композиция:

$$\mu_B(y) = \max_{x \in X} \left\{ \left\{ \mu_A(x), \mu_Q(\langle x, y \rangle) \right\} \right\}.$$

Min-max-композиция:

$$\mu_B(y) = \min_{x \in X} \left\{ \max \left\{ \mu_A(x), \mu_Q(\langle x, y \rangle) \right\} \right\}.$$

Max-max-композиция:

$$\mu_B(y) = \max_{x \in X} \left\{ \max \left\{ \mu_A(x), \mu_Q(\langle x, y \rangle) \right\} \right\}.$$

Min-min-композиция:

$$\mu_B(y) = \min_{x \in X} \left\{ \min \left\{ \mu_A(x), \mu_Q(\langle x, y \rangle) \right\} \right\}.$$

Max-average-композиция:

$$\mu_B(y) = 0,5 \cdot \max_{x \in X} \left\{ \mu_A(x) + \mu_Q(\langle x, y \rangle) \right\}.$$

Sum-prod-композиция:

$$\mu_B(y) = f \left(\sum_{x \in X} (\mu_A(x) \cdot \mu_Q(\langle x, y \rangle)) \right).$$

Результаты применения разработанной методики

Рассмотренная методика реализована в программном средстве «Автоматизированная информационная система составления технического задания с экспертной оценкой принятия решения», предназначенном для поддержки руководителя программного проекта при разработке технического задания на ПО с подбором альтернативного варианта интерфейса [12]. Главное окно программного средства показано на рисунке 2.

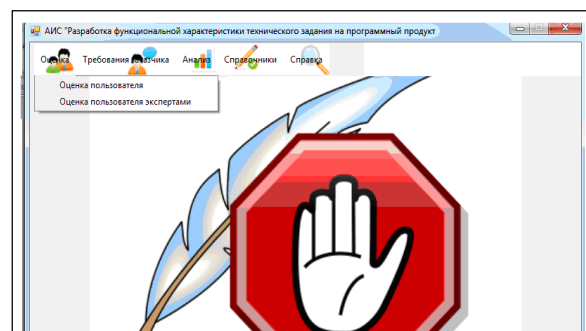


Рис. 2. Главное окно программного средства

Fig. 2. The software tool main screen

Одной из функций разработанного программного средства является тестирование пользователя, реализованное в форме «Оценка пользователем», в которой можно выбрать характеристику и ее градацию из выпадающих меню (http://www.swsys.ru/uploaded/image/2017_1/2017-1-dop/5.jpg).

Оценка пользователя экспертами (выбор его характеристик и их градаций) основывается на результатах тестирования пользователя и проводится

в форме «Оценка пользователя экспертами» (http://www.swsys.ru/uploaded/image/2017_1/2017-1-dop/6.jpg).

Результаты вычислений выводятся в окне «Анализ» (http://www.swsys.ru/uploaded/image/2017_1/2017-1-dop/7.jpg). По нажатии кнопки «Сохранить отчет» генерируется документ, в котором заполнены следующие разделы: наименование ПО, материально ответственное лицо (руководитель проекта), наименование заказчика, характеристики пользователя, рекомендуемый шаблон интерфейса, назначение и область применения ПО, требования к функциональным характеристикам ПО, требования к составу и параметрам технических средств, требования к информационной и программной совместимости, состав программной документации, стадии, этапы разработки и сроки сдачи ПО.

Во вкладке «Справочники» обеспечивается работа с информацией: должности, тип юридического лица, тип предметной области, тип характеристик пользователя, тип шаблона, физическое лицо, группа пользователей, единицы измерения. Эту информацию можно корректировать, добавлять и удалять.

Заключение

Разработанное программное средство ведет базу знаний о различных категориях пользователей с учетом их компьютерной грамотности, моторики, памяти, мышления, дальтонизма, концентрации внимания и др. На основе этих характеристик разрабатываются различные варианты пользовательского интерфейса, которые согласуются с заказчиком и реализуются в конечном ПО.

Основными результатами работы является программное и математическое обеспечение для решения задач интеллектуального проектирования пользовательского интерфейса ПО с учетом индивидуальных характеристик пользователей.

Решены следующие научно-практические задачи:

- выявлены основные параметры, по которым можно классифицировать пользователей при разработке интерфейсов ПО;

- разработана методика проектирования адаптированных пользовательских интерфейсов с элементами искусственного интеллекта;

- разработана нечеткая модель на бинарных нечетких отношениях для подбора альтернативного шаблона интерфейса ПО.

Предложенная методика позволит автоматизировать процесс проектирования пользовательского интерфейса, конкретизировать техническое задание на ПО, что в конечном итоге повысит объективность и оперативность решений, принимаемых разработчиками.

Литература

1. Брауде Э. Технология разработки программного обеспечения. СПб: Питер, 2004. 655 с.
2. Кулямин В.В. Технологии программирования. Компонентный подход. М.: ИНТУИТ; БИНОМ. Лаборатория знаний, 2011. 463 с.
3. Таганов А.И., Гильман Д.В. Методологические основы анализа и аттестации уровней зрелости процессов программных проектов в условиях нечеткости: монография. М.: Горячая линия-Телеком, 2014. 168 с.
4. Ишакова Е.Н., Зубкова Т.М. Автоматизация управления рисками программных проектов на основе нечеткого логического вывода // Науч.-технич. вестн. информ. технологий, механики и оптики. 2015. № 5. С. 877–885.
5. Kudinov Yu.I. Synthesis of a fuzzy-logic control system. Jour. of Comp. and Syst. Sc. Intern. 1999, no. 1, pp. 158–164.
6. Chen J., Rine D.C. Training fuzzy logic controller software components by combining adaptation algorithms. Advances in Engineering Software. 2003, no. 3, pp. 125–137.
7. Kandel A., Zhang Y.-Q., Henne M. On use of fuzzy logic technology in operating systems. Fuzzy Sets And Systems. 1998, no. 3, pp. 241–251.
8. Wang J. A fuzzy project scheduling approach to minimize schedule risk for product development. Fuzzy Sets And Systems. 2002, no. 2, pp. 99–116.
9. Carr V., Tah J.H.M. A fuzzy approach to construction project risk assessment and analysis: construction project risk management system. Advances in Engineering Software. 2001, no. 10–11, pp. 847–857.
10. Джарратано Д., Райли Г. Экспертные системы: принцип разработки и программирование. 4-е изд. М.: Вильямс, 2007. 1152 с.
11. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. СПб: БХВ-Петербург, 2005. 736 с.
12. Зубкова Т.М., Колобов А.Н. Автоматизированная информационная система составления технического задания с экспертной оценкой принятия решения. Свид. о гос. регистр. прог. для ЭВМ № 2016616452. Заявл. 18.04.2016. Опубл. 10.06.2016.

SOFTWARE INTERFACE DESIGN USING ELEMENTS OF ARTIFICIAL INTELLIGENCE

T.M. Zubkova¹, Dr.Sc. (Engineering), Professor, bars87@mail.ru

E.N. Natochaya², Ph.D. (Education), Associate Professor, en_ischa@mail.ru

¹ Orenburg State University, Pobedy Ave. 13, Orenburg, 460018, Russian Federation

² Russian Presidential Academy of National Economy and Public Administration (Orenburg branch), Curacha St. 26, Orenburg, 460000, Russian Federation

Abstract. In order to develop high-quality software it is necessary to reflect all customer requirements in the specification, thus, to have a global view on the future software for customers and performers. One of the options to achieve mutual understanding is to develop a prototype of a user interface.

The article describes the methods of selecting an alternative version of the interface template using such artificial intelligence methods as expert evaluation and the fuzzy-set theory. Users might be divided into five groups on the basis of individual characteristics (a newbie, usual, experienced, skilled, an administrator). The article defines the basic parameters of individual characteristics which may help to classify users when designing interfaces (computer literacy, systematic experience, experience of working with similar programs, typing, thinking, memory, motor skills, blindness, concentration, emotional stability).

The paper describes mathematical support and software for solving the problems of intelligent user interface design. Task implementation is performed in three stages. The first stage is "Forming and assessing expert group competence". It defines the characteristics of experts. A quantitative description of experts' characteristics is based on the calculation of relative ratios of competence according to the results of experts' statements on the Advisory group. The second stage is "Group expert assessment of the object with direct assessment". It determines recurrent relations for iterations. The third phase is "Building a fuzzy model on fuzzy binary relations". It operates by two fuzzy sets: a set of user groups and a variety of interface templates that are maximally effective for users with these characteristics. Fuzzy model input data are selected fuzzy sets, the output data are the degrees of matching interface patterns to users.

The user interface design process is automated on the basis of the proposed methodology in order to improve objectivity and optimize decisions taken by software developers.

Keywords: software, user interface, template interface, interface prototype, technical specification, expert assessment, knowledge base, fuzzy relations.

References

1. Braude E. *Tekhnologiya razrabotki programmnogo obespecheniya* [The Software Development Technology]. St. Petersburg, Piter Publ., 2004, 655 p.
2. Kulyamin V.V. *Tekhnologii programmirovaniya* [Programming Technologies. The component approach]. Moscow, Internet-University of Information Technology, BINOM. Laboratoriya Znaniy Publ., 2011, 463 p.
3. Taganov A.I., Gilman D.V. *Metodologicheskie osnovy analiza i attestatsii urovney zrelosti protsessov programmykh proektov v usloviyakh nechetkosti* [Methodological framework for analysis and assessment of software project process maturity levels in terms of fuzziness]. Monograph, Moscow, Goryachaya liniya-Telecom Publ., 2014, 168 p.
4. Ishakova E.N., Zubkova T.M. Risk management automation of software projects based on fuzzy inference. *Nauch.-tekhnich. vestn. inform. tekhnology, mekhaniki i optiki* [Scientific and Technical Jour. of Information Technologies, Mechanics and Optics]. 2015, no. 5, pp. 877–885 (in Russ.).
5. Kudinov Yu.I. Synthesis of a fuzzy-logic control system. *Jour. of Computer and Systems Sciences Int.* 1999, no. 1, pp. 158–164.
6. Chen J., Rine D.C. Training fuzzy logic controller software components by combining adaptation algorithms. *Advances in Engineering Software*. 2003, no. 3, pp. 125–137.
7. Kandel A., Zhang Y.-Q., Henne M. On use of fuzzy logic technology in operating systems. *Fuzzy Sets And Systems*. 1998, no. 3, pp. 241–251.
8. Wang J. A fuzzy project scheduling approach to minimize schedule risk for product development. *Fuzzy Sets And Systems*. 2002, no. 2, pp. 99–116.
9. Carr V., Tah J.H.M. A fuzzy approach to construction project risk assessment and analysis: construction project risk management system. *Advances in Engineering Software*. 2001, no. 10–11, pp. 847–857.
10. Giarratano J.C., Riley G.D. *Expert systems: the principle of development and programming*. 4th ed. Course Technology Publ., 2004, 288 p. (Russ.ed.: Moscow, Vilyams Publ., 2007, 1156 p.).
11. Leonenkov A.V. *Nechetkoe modelirovanie v srede MATLAB i fuzzyTECH* [Fuzzy modeling in MATLAB and fuzzyTECH]. St. Petersburg, BHV-Peterburg Publ., 2005, 736 p.
12. Zubkova T.M., Kolobov A.N. *Avtomatizirovannaya informatsionnaya sistema sostavleniya tekhnicheskogo zadaniya s ekspertnoy otsenkoy prinyatiya resheniya* [Automated Information System for Creating Technical Task with Expert Assessment of Decision-Making]. Software State Registration Certificate no. 2016616452. 2016.

Примеры библиографического описания статьи

1. Zubkova T.M., Natochaya E.N. Проектирование интерфейса программного обеспечения с использованием элементов искусственного интеллекта // Программные продукты и системы. 2017. Т. 30. № 1. С. 5–11; DOI: 10.15827/0236-235X.117.005-011.

2. Zubkova T.M., Natochaya E.N. Software interface design using elements of artificial intelligence. *Programmnye produkty i sistemy* [Software & Systems]. 2017, vol. 30, no. 1, pp. 5–11 (in Russ.); DOI: 10.15827/0236-235X.117.005-011.

УДК 004.65, 004.032.26
DOI: 10.15827/0236-235X.117.012-020

Дата подачи статьи: 11.08.16
2017. Т. 30. № 1. С. 12–20

РАЗРАБОТКА СИСТЕМЫ ХРАНЕНИЯ АНСАМБЛЕЙ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ

Е.В. Пучков, к.т.н., доцент, puchkoff@i-intellect.ru;

С. Терехов, магистрант, isergeiterehov@gmail.com

(Академия строительства и архитектуры

Донского государственного технического университета,

ул. Социалистическая, 162, г. Ростов-на-Дону, 344022, Россия)

Важным инструментом в работе специалиста по анализу данных и машинному обучению является ПО для организации экспериментов. Прежде всего это связано с большим количеством этапов в обработке данных и спецификой их выполнения. В ходе работы был спроектирован и разработан прототип системы хранения ансамблей нейросетевых моделей, обеспечивающий структурированное хранение данных на различных этапах решения задачи прогнозирования временных рядов. Рассмотрены модель данных, архитектура системы хранения и механизмы поступления и перераспределения информации в ней. Разработана модель классов для программного взаимодействия с хранилищем. Для хранения данных об объектах и связей между этими объектами была использована MySQL, а для хранения временных рядов – нереляционная БД InfluxDB. Создан пользовательский интерфейс с возможностями наглядного отображения данных и удобного взаимодействия с хранилищем ансамблей нейросетевых моделей. Апробация системы проводилась на примере задачи прогнозирования солнечной активности за период с января 1700 года по февраль 2015 года. Проведенный эксперимент с применением рекуррентной сети LSTM показал, что ошибка ансамбля нейросетевых моделей ниже ошибки каждой отдельно взятой нейросетевой модели. LSTM построена с применением библиотеки Keras, для формирования ансамбля использован подход Blending.

Результаты проделанной работы показывают перспективность разработки, обеспечивающей высокую степень интеграции в расширяемые программные продукты на языке Python. Разработка полнофункциональной системы позволит не только организовать процесс анализа данных, но и повысить качество результирующих моделей за счет автоматизации процесса формирования ансамблей.

Ключевые слова: хранилище, нереляционная БД, рекуррентные нейронные сети, LSTM, ансамбль, stacking, прогнозирование временных рядов.

В последнее время среди специалистов по анализу данных и машинному обучению все более популярным становится ПО для организации исследований. Прежде всего это связано с большим количеством этапов обработки данных и спецификой их выполнения. Можно выделить такую библиотеку, как Sacred [1], которая позволяет организовать эксперименты без привязки к конкретным моделям, данные параметров моделей и результаты можно сохранить в БД. В библиотеке Huperopt [2] акцент делается на оптимизации параметров моделей. FGLab [3] позволяет аналитику запускать свои модели на распределенной системе с возможностью сохранять результаты экспериментов и их параметры в БД. Для сложных вычислительных задач с применением Nadoor, которые могут длиться дни или недели, подойдет Luigi [4]. Данный пакет позволит организовать управление многочисленными вычислительными задачами в одном месте. Последние две системы имеют интерфейс для визуализации результатов и информации по задачам.

Заключительным этапом в решении задачи машинного обучения является построение ансамбля моделей, поскольку в некоторых случаях оптимальное решение может быть получено с применением ансамбля нескольких различных моделей. Большое количество источников показывают практическую значимость применения ансамбля в решении прикладных задач [5–7]. Очень часто в таких ансамблях используют нейросетевые модели.

Примечательно, что построение ансамбля только из нейросетевых моделей в некоторых задачах даст преимущество [8–10]. В связи с этим возникает проблема хранения данных на этапах моделирования, в том числе данных самих моделей и построенных с их помощью ансамблей. Проведенный обзор систем организации экспериментов показал, что существующие системы не решают такую проблему в явном виде.

Цель данной работы – проектирование и разработка системы хранения ансамблей нейросетевых моделей, обеспечивающей структурированное хранение данных на различных этапах решения задач прогнозирования временных рядов. Разработка хранилища позволит не только организовать процесс анализа данных, но и повысить качество результирующих моделей за счет автоматизации процесса формирования ансамблей.

Работу можно разделить на следующие основные части:

- разработка модели данных и архитектуры системы хранения;
- разработка пользовательского интерфейса;
- тестирование системы на реальных данных.

Для задач прогнозирования временных рядов принято использовать два типа ИНС: рекуррентные сети (RNN) [11] и различные сети прямого пространства с задержкой по времени (TLFN) [12]. Задержку по времени также можно применять и для рекуррентных сетей [13]. В работе при реше-

нии задачи прогнозирования временного ряда будет использована LSTM (long short-term memory – долгая краткосрочная память). Рекуррентные нейронные сети, основанные на этом подходе, получили большое распространение при решении задач распознавания рукописного текста, моделирования языка, машинного перевода, обработки аудио- и видеоизображений, анализа тональности и классификации текстов, прогнозирования временных рядов.

При решении сложных задач классификации, регрессии, а также прогнозирования временных рядов часто оказывается, что ни один из алгоритмов не обеспечивает желаемого качества восстановления зависимости. В таких случаях имеет смысл строить композиции алгоритмов (ансамбли), в которых ошибки отдельных алгоритмов взаимно компенсируются. Для задачи прогнозирования временных рядов подойдут такие подходы, как голосование и стекинг (stacking). Они подразумевают формирование ансамбля из моделей, полученных на одинаковых данных, что подходит для временных рядов, в отличие от бустинга (boosting) и бэггинга (bagging), где для базовых алгоритмов используются разные данные.

Наиболее известные корректирующие операции голосования:

– простое:

$$b(x) = F(b_1(x), \dots, b_T(x)) = \frac{1}{T} \sum_{t=1}^T b_t(x);$$

– взвешенное:

$$b(x) = F(b_1(x), \dots, b_T(x)) = \sum_{t=1}^T w_t b_t(x), \quad \sum_{t=1}^T w_t b_t(x), \quad \sum_{t=1}^T w_t = 1, \quad w_t \geq 0;$$

– смесь экспертов:

$$b(x) = F(b_1(x), \dots, b_T(x)) = \sum_{t=1}^T w_t(x) b_t(x), \quad \sum_{t=1}^T w_t = 1, \quad \forall x \in X.$$

Простое голосование – это лишь частный случай взвешенного голосования, а взвешенное является частным случаем смеси экспертов.

Основная идея стекинга и его разновидности блендинга заключается в использовании базовых алгоритмов для получения предсказаний (метапризнаков) и использовании их как признаков для некоторого обобщающего алгоритма (метаалгоритма). Иными словами, основной идеей стекинга является преобразование исходного пространства признаков задачи в новое пространство, точками которого являются предсказания базовых алгоритмов [14].

Разработка модели данных и архитектуры системы хранения

Для реализации поставленных задач необходим следующий набор программных средств:

- реляционная БД для хранения данных об объектах и связей между этими объектами;
- нереляционная БД для хранения временных рядов;
- язык программирования для реализации логики системы хранения;
- сопутствующие программные пакеты, в том числе реализующие LSTM.

В качестве реляционной СУБД была использована MySQL [15]. Для хранения временных рядов современное решение – InfluxDB [16]. Для программирования логики хранилища нейросетевых моделей выбраны Python версии 2.7.11 и следующие свободно распространяемые пакеты:

- numpy (для работы с массивами [17]);
- sklearn (библиотека для анализа данных [18]);
- cherrypy (библиотека, позволяющая реализовать веб-сервер [19]).

Среди многочисленных программных реализаций архитектур нейронных сетей, в частности LSTM, выделим Theano [20], а также созданную на ее основе библиотеку Keras [21]. Библиотека Keras позволяет использовать как Theano, так и TensorFlow [22] в качестве основы вычислений. Keras упрощает процесс создания нейронных сетей, предоставляя для этого специальный конструктор. В основе любого кода с использованием Keras лежит объект model, который описывает то, в каком порядке и какие именно слои содержит ваша нейронная сеть.

Для построения структуры реляционной БД рассмотрим необходимые сущности и их структуру [23].

Проект (project) объединяет ряд исследований над набором данных. В рамках проекта рассматриваются данные из определенного источника (временной ряд, который хранится в InfluxDB). Все действия по преобразованию данных, построению моделей или ансамблей производятся в рамках проекта.

Источник данных (data_source) представляет собой описание данных в источнике. В рамках хранилища рассматривается основная задача – прогнозирование временных рядов. Соответственно, информация об источнике данных включает такую информацию, как начало периода, конец периода, интервал измерений и другие. В связи с тем, что источник данных не фиксирован, то есть данные в нем могут изменяться, дополняться, удаляться, необходимо фиксировать состояние источника данных на момент начала какого-либо исследования или ряда исследований.

Снимок данных (data_snapshot) отражает состояние источника данных на момент времени. Однако сами данные в исходном виде, как правило, не пригодны для построения качественных моделей, поэтому необходимо выполнить ряд преобразований.

Преобразование данных (`data_preparation`) показывает способ преобразования данных (снимка данных), а также сохраняет преобразованные данные для дальнейшего применения. Преобразованные данные по-прежнему являются временным рядом, но для использования в различных нейросетях должны быть созданы конечные наборы данных в виде матриц X и Y , объясняющие признаки и целевые значения.

Набор данных (`data_set`) – это конечная выборка данных, отвечающая требованиям той или иной модели. Например, одна модель может использовать для прогнозирования окно в 7 значений, а целевое (прогнозируемое) значение будет отступать на 2 пункта от окна. В этом случае размерность матрицы X – 7, а Y формируется по определенному правилу. При других параметрах выборка будет сформирована иначе, что и объясняет необходимость введения рассматриваемой сущности.

Снимок данных, преобразование данных и набор данных – это отдельные наборы данных, пошагово полученные из предыдущего источника. Эти данные уже необязательно являются временными рядами с точки зрения способа их хранения. В связи с этим необходимо организовать хранение этих данных в унифицированном виде. Наиболее подходящим форматом хранения является CSV (Comma-Separated Values – разделяемые запятыми значения) – текстовый формат, предназначенный для представления табличных данных. Определим также сущность CSV-данных (`data_csv`) – это зависимая сущность, которая представляет собой только сами данные под уникальным идентификатором.

Другим не менее важным набором сущностей является набор, связанный с нейросетевыми моделями.

Исследование (`research`) – это группа моделей, полученных по определенным правилам. Такие правила устанавливают порядок преобразования данных, способ построения моделей и их настройки и т.д. В рамках исследования рассматриваются данные, преобразованные определенным образом, поэтому все модели, построенные в ходе исследования, с точки зрения представления работают с одними и теми же данными.

Модель (`model`) – это представление математической модели. Модель может быть любого типа: как нейросетевой, так и любой другой (случайный лес, логистическая регрессия и др.). Для получения модели данные должны быть подготовлены определенным образом, как говорилось ранее, и сохранены как набор данных. Такой набор и используется далее для обучения и тестирования модели.

В ходе описания настроек или построения модели могут возникать некоторые данные, так или иначе описывающие модель. Они называются метаданными и требуют вынесения в определенную сущность (для реляционной БД).

Метаданные модели (`model_meta`) – это простое представление данных о модели в виде «ключ-значение». Такие данные могут содержать настройки модели и/или данные о процессе обучения (ошибка, доля правильных ответов, время обучения, алгоритм обучения и другие).

Последним набором сущностей являются сущности, связанные с ансамблями (комитетами).

Ансамбль (`ensemble`) – это сущность, создаваемая в рамках одного проекта. Такое ограничение вводится для ограничения данных: все модели должны работать на данных одного и того же рода. Здесь описываются такие данные об ансамбле, как метод построения ансамбля, тип метамодели (метаклассификатора), математическая модель ансамбля и т.д.

Ансамбль составляется из моделей. При этом каждая модель может содержать ряд определенных параметров с точки зрения ансамбля, например, вес эксперта для линейной регрессии.

Элемент ансамбля (`ensemble_item`) – параметризованная модель, используемая в построении ансамбля.

Физическая модель данных в MySQL представлена на рисунке 1. Для связи реляционной БД MySQL и нереляционной InfluxDB необходимо ввести ряд спецификаций:

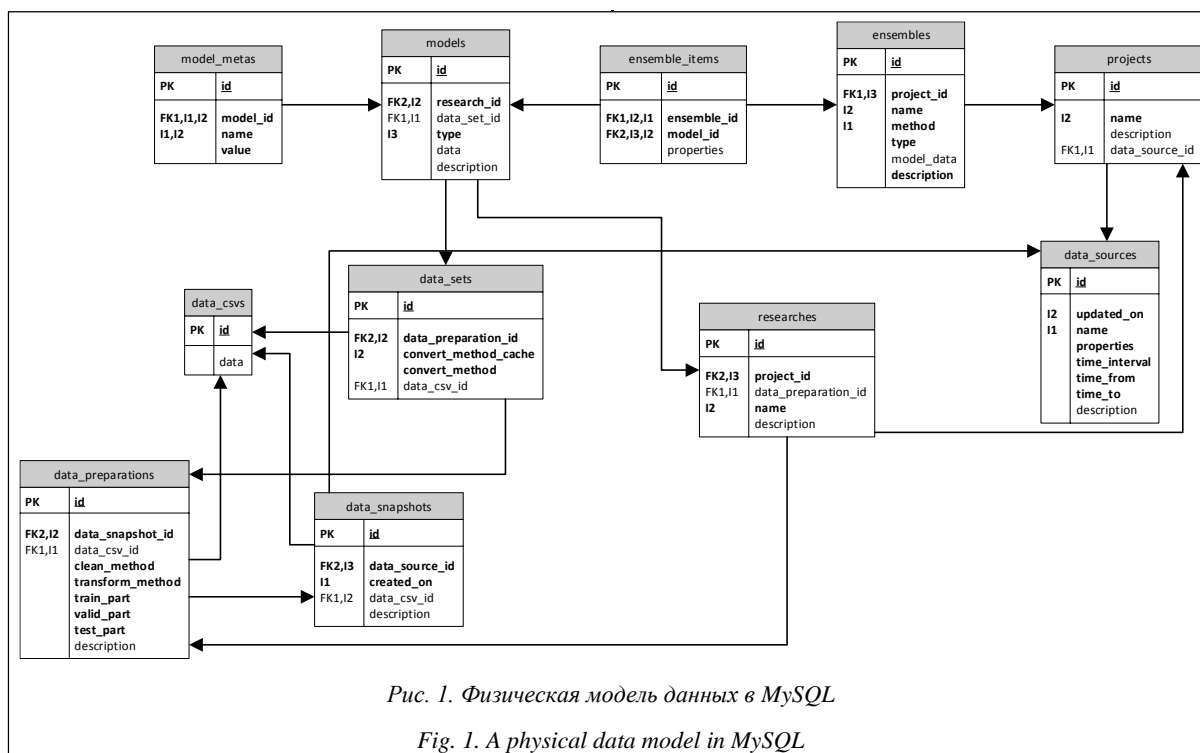
- измерение (`measurement`), содержащее экспортируемую информацию, должно иметь имя `data_source`;
- измерение обязательно должно включать тэг (`tag`) `mysql_id`, содержащий идентификатор исходных данных, куда будет произведена привязка;
- тип данных значения (`value`) должен быть `float`-числом с плавающей точкой.

Измерение создается автоматически при добавлении новых данных. Рассмотрим пример добавления данных в необходимое измерение по установленным правилам: `data_source,mysql_id=234 value=0.55 1422568543702900257`. Такой запрос добавит в БД InfluxDB запись в необходимое измерение для источника данных с идентификатором 234. Запись будет содержать значение 0.55 и привязку ко времени со значением 1422568543702900257 (`timestamp`) – Thu, 29 Jan 2015 21:55:43.702900257 GMT.

Одна из основных задач хранилища – предоставление функционального программного интерфейса для взаимодействия с данными, хранящимися в БД. Поэтому, помимо средств хранения данных (MySQL, InfluxDB), хранилище ансамблей нейросетевых моделей включает внутреннюю логику, определяющую правила функционирования системы.

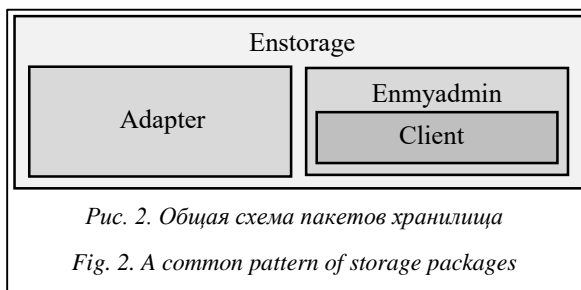
Рассмотрим каждый пакет из представленных на рисунке 2:

- `enstorage` (полная библиотека хранилища, включающая в себя основные модели объектов, отражающие сущности БД (ORM, Object-Relational Mapping));



– adapter (библиотека, реализующая методы преобразования данных для дальнейшего использования в моделях);

– enmyadmin (содержит основной функционал встроенной системы администрирования хранилища).



Отдельным классом, который обязательно должен быть использован перед работой с хранилищем ансамблей нейросетевых моделей, является Connector (рис. 3). Он осуществляет подключение к необходимым БД (MySQL и InfluxDB). Все остальные классы являются компонентами ORM и реализуют следующие стандартные public (доступные извне) функции:

- delete – удаление связанного с БД объекта;
- save – сохранение (создание или обновление) объекта БД;
- get – статичный метод, возвращающий объект класса, к которому он относится, по указанному идентификатору записи (id);
- get_list – статичный метод, возвращающий список объектов класса, в котором вызван, по указанным идентификаторам (ids); если идентифика-

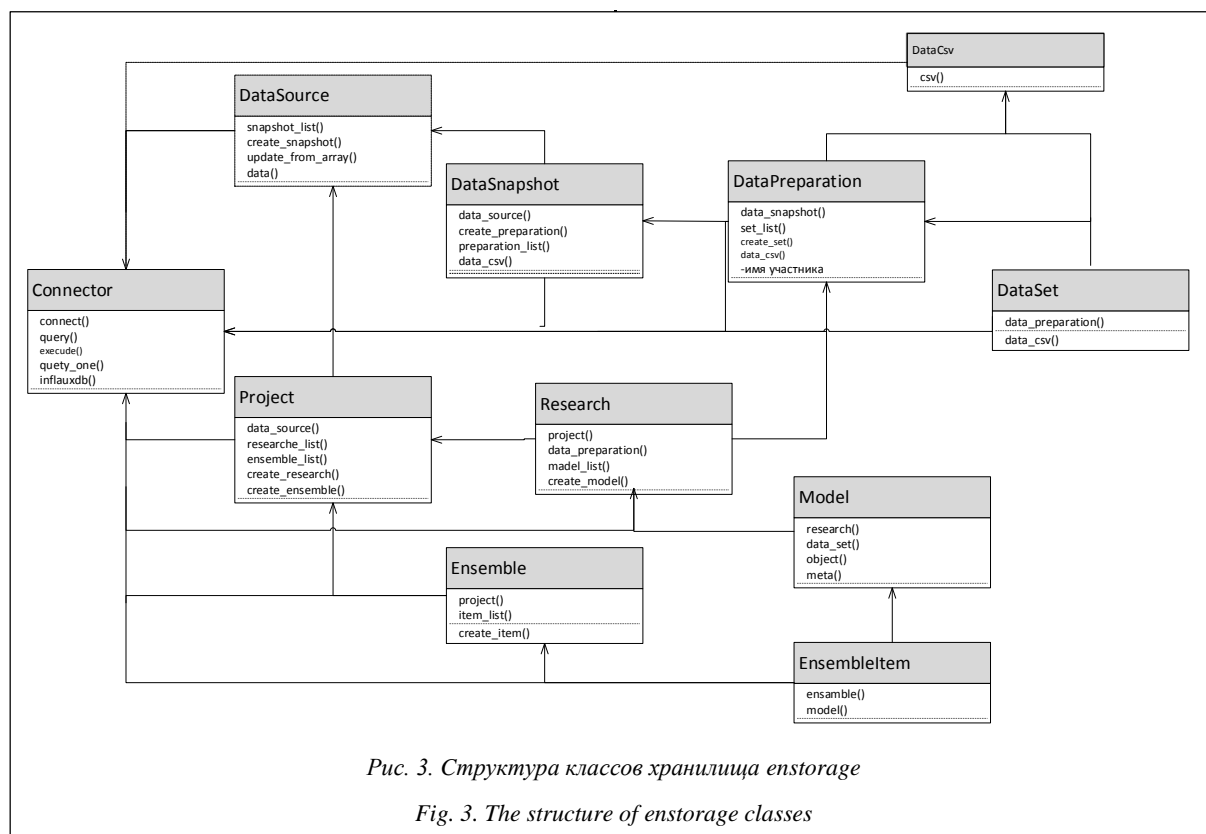
торы не указаны, возвращается полный список всех объектов.

С течением времени исходные данные в InfluxDB могут обновляться и пополняться. Для фиксации определенного набора данных необходимо создать снимок (DataSnapshot) (рис. 4).

Снимок создается при помощи метода DataSource.create_snapshot(), который загружает текущее состояние источника на указанный временной период (DataSource: time_from, time_to). Загрузка данных выполняется с применением агрегирующей функции, группирующей данные по временному интервалу, – DataSource.time_interval. Сохраненный снимок выступает в роли самостоятельных данных, которые могут быть использованы для дальнейших исследований.

Для подготовки данных создается объект DataPreparation, который обеспечивает хранение подготовленных данных, а также преобразование данных DataSnapshot. Создание преобразованных данных выполняется при помощи метода DataSnapshot.create_preparation(clean_method, transform_method, train_part, valid_part, test_part), где параметрами являются (по порядку) метод заполнения пустых значений, метод преобразования значений, доля обучающей выборки, доля валидационной выборки, доля тестовой выборки. Подготовка данных в ручном режиме осуществляется методом DataPreparation.prepare().

После преобразования данных на их основе может быть создано исследование – Research. При создании модели Research не определяет, с каким набором работает модель, это задача сервиса, использующего хранилище.



Сервис должен произвести следующие действия:

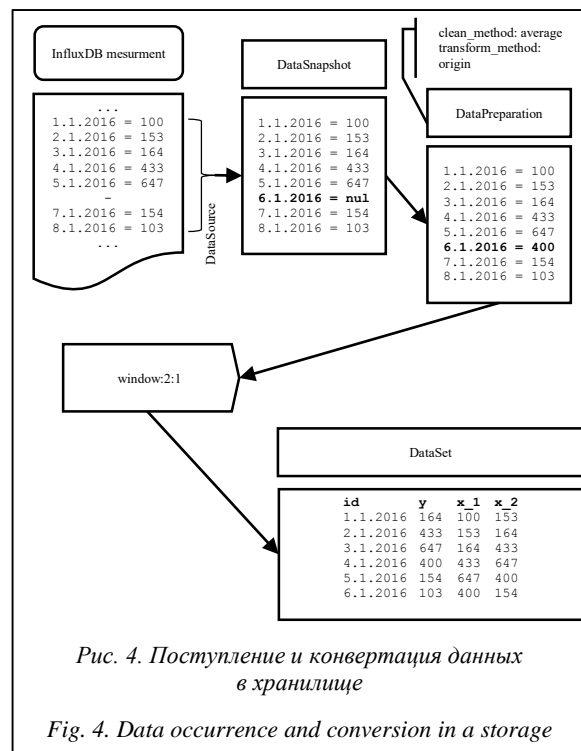
- создать модель, привязанную к исследованию;
- исходя из типа модели преобразовать нужным адаптером (adapter) данные – `DataPreparation.create_set(adapter)`;
- сообщить модели о созданном наборе данных посредством `Model.data_set(created_data_set)`.

При этом в автоматическом режиме адаптером будут обработаны преобразованные данные `DataPreparation`. Данный этап является завершающим для серии преобразования исходных данных, полученных из `DataSource`.

Еще одним важным потоком данных является информация, поступающая в ходе построения моделей и ансамблей. Фиксация таких данных (метаданных) для модели осуществляется методами `Model.meta(key, value)`. Метаданными могут быть абсолютно любые данные, описывающие модель. При построении ансамбля дополнительные параметры фиксируются в свободной форме в `EnsembleItem.properties`, однако рекомендуется использовать формат JSON.

Особенностью хранилища является то, что хранение объектов конечных реализаций моделей осуществляется благодаря специальному формату. Данные объектов сериализуются и десериализуются при помощи библиотеки `pickle`. Такой подход обеспечивает возможность сохранения и восстановления объектов целиком, тем самым обеспечи-

вая высокий уровень интеграции пакета `enstorage` с другими библиотеками. Также, благодаря используемому формату, хранилище может принять не только нейросетевые модели, но и любые другие. Однако из-за использования библиотеки `pickle` существует ограничение на использование этих дан-



ных в языках, отличных от Python, так как данные совместимы только с ним.

Разработка пользовательского интерфейса

Для упрощения администрирования хранилища ансамблей нейросетевых решений предусмотрен пакет `enmyadmin`, входящий в `enstorage`. Данный пакет представляет собой веб-сервер с основными методами администрирования. Основным шаблоном проектирования веб-приложения является `model-view-controller` (MVC, «модель-представление-контроллер»). В роли клиентского приложения выступает HTML5-JS-приложение, разработанное с использованием AngularJS [24]. Взаимодействие клиентского и серверного приложений осуществляется по технологии REST (representational state transfer – «передача состояния представления»), обеспечивающей независимость серверной части от клиентского приложения. Фреймворк работает с HTML, включающим дополнительные пользовательские атрибуты, которые описываются директивами, и связывает ввод-вывод области страницы с моделью, состоящей из объектов JavaScript. Значения этих объектов задаются вручную или извлекаются из статических или динамических JSON-данных.

На рисунке 5 представлена форма просмотра информации о проекте. В левой части формы находится фиксированная панель навигации, позволяющая просматривать список проектов и источников данных для быстрого перехода к ним. Для поиска необходимого пункта предусмотрен функционал фильтрации. В правой части окна находится область управления, включающая элементы управления открытым объектом.

Разработаны следующие формы управления объектами: проект, исследование, модель, ансамбль, исходные данные, снимок данных, преобразованные данные, набор данных.

Важно отметить, что благодаря использованию технологии REST клиентское приложение системы администрирования может быть разработано на любой платформе, поддерживающей взаимодействие по HTTP-протоколу.

Тестирование системы на реальных данных

Для тестирования работоспособности хранилища в реальных задачах необходимо реализовать систему построения ансамблей, а также обучения моделей. Конструктор ансамблей – это отдельный функционал, который может быть вынесен в специальный пакет `endirector`. Данный пакет включает методы построения ансамбля и использует объекты хранилища из пакета `enstorage`. Ансамбль формируется в автоматическом режиме на основе настроек в `Ensemble`. Далее приведен пример построения ансамбля из набора моделей, а также применения `Conductor` для формирования метамодели:

```
project = enstorage.Project.get(1)
models = enstorage.Model.get_list([3,4,5])
# Настройка ансамбля
ensemble = project.create_ensemble('blending', 'LinearRegression', models=models)
# Создание ансамбля
conductor = endirector.Conductor(ensemble)
conductor.render()
```

Порядок действий для построения ансамбля, реализованный в методе `Conductor.render()`:

- инициализация объекта метамодели;
- определение метода создания ансамбля;
- создание пересечения матриц данных (DataSet) всех экспертов;

EnMyAdmin

Навигация

Проекты

Создать

Фильтр

· Solar activity (35)

Количество 1 из 1

Источники данных

Создать

Фильтр

· Test data source (1)

Количество 1 из 1

Управление

Проект “Solar activity”

Основные свойства

id	35
Имя	Solar activity
Исходные данные	Test data source

Описание

Testing the system

Действия

Удалить

Список ансамблей

id	Имя	Тип	Метод	Описание
26	Simple ensemble	LinearRegression	Blending	Testing

Список исследований

id	Имя	Описание
86	Standard data	Zoom on 0.1

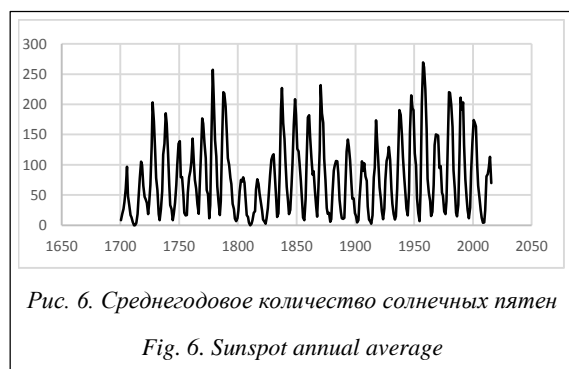
Рис. 5. Клиентское HTML-приложение, реализующее интерфейс EnMyAdmin

Fig. 5. Client HTML-application that implements EnMyAdmin interface

- запуск алгоритма построения ансамбля;
- сохранение метамодели и параметров моделей.

Стоит отметить, что наборы данных различных моделей могут существенно отличаться друг от друга. Так, один набор данных может быть получен с задержкой в 4 значения, а другой – в 10. При этом объем выборок также будет отличаться. Могут использоваться и другие методы конвертации временного ряда, что делает невозможным однозначное получение результата всех моделей для одних данных. Решением данной проблемы является идентификация (*id*) целевых значений. Таким образом, каждый набор $[X, y]$ в DataSet также включает и *id* – $[id, X, y]$. Благодаря этому можно получить значения всех моделей для конкретного целевого значения – *y*.

В качестве исходных данных для тестирования системы использованы данные о солнечной активности за период с января 1700 года по февраль 2015 года, всего 303 значения (рис. 6). Для эксперимента построим 3 нейронные сети с задержкой в 5, 7, 13 значений.



В ходе выполнения итогового скрипта осуществляются следующие действия:

- подключение к хранилищу;
- получение данных о наборе данных;
- создание снимка данных;
- преобразование данных (масштабирование, выделение тестовой выборки – 30 %);
- создание проекта и исследования;
- создание и инициализация модели;
- обучение моделей;
- создание ансамбля.

Далее приведен код создания модели LSTM с применением библиотеки Keras:

```
model = Sequential()
model.add(LSTM(output_dim=50, input_dim=n_input,
return_sequences=True))
model.add(Dropout(0.2))
model.add(LSTM(100, return_sequences=False))
model.add(Dropout(0.2))
model.add(Dense(output_dim=1))
model.add(Activation("linear"))
model.compile(loss=item.meta('loss'), optimizer='rmsprop')
```

Для оценки качества каждой модели, а также ансамбля рассчитаем среднеквадратическую ошибку (MSE) на тестовой выборке.

На рисунке 7 видим, что наименьшее значение ошибки у ансамбля (на графике – out). Так как в качестве метамодели использовалась линейная регрессия (Linear Regression), вес каждой модели можно оценить в результирующем значении. Данные значения записаны в EnsembleItem.properties: 1.06192747 – keras-1, -0.13015426 – keras-2, 0.17757505 – keras-3.

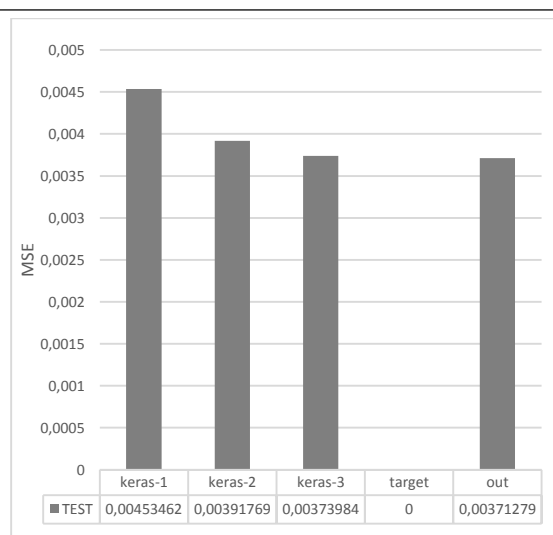


Рис. 7. Среднеквадратическая ошибка моделей и ансамбля на тестовой выборке

Fig. 7. Mean square error of models and an ensemble on a test set

Значения весов можно интерпретировать следующим образом: наибольшим весом обладает первая модель (keras-1), небольшую корректировку вносит третья модель (keras-3), компенсацию оказывает вторая модель (keras-2). На рисунке 8 представлены результаты прогноза, полученные с помощью ансамбля.

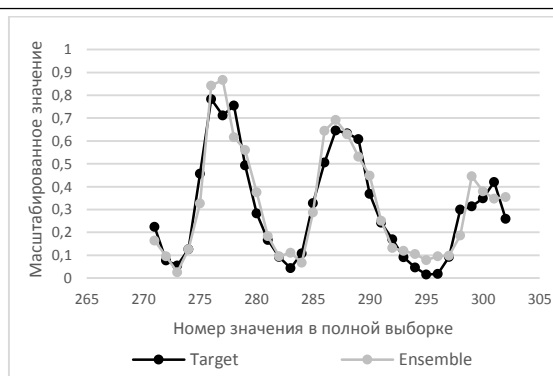


Рис. 8. График прогноза солнечной активности на тестовой выборке

Fig. 8. A solar activity forecast diagram on a test set

Заключение

В ходе выполнения данной работы был создан прототип системы хранения ансамблей нейросетевых моделей. Проведенный эксперимент по прогнозированию солнечной активности показал, что ошибка ансамбля нейросетевых моделей ниже ошибки каждой отдельно взятой нейросетевой модели. Несомненно, для улучшения результатов прогнозирования необходимы дополнительные эксперименты и совершенствование ПО.

Разработаны следующие программные решения:

- пакет для языка Python, обеспечивающий быстрое и упрощенное взаимодействие с БД, реализованный с использованием технологии ORM;
- пакет преобразования временного ряда в конечные выборки, применяемые в моделях;
- интерфейс пользователя в виде HTML-приложения, обеспечивающий наглядное отображение данных и удобное взаимодействие с хранилищем ансамблей нейросетевых моделей.

Результаты проделанной работы показывают перспективность разработанных программных решений и обеспечивают высокую степень интеграции в расширяемые программные продукты на языке Python.

Исследование выполнено при финансовой поддержке РФФИ, проект № 14-01-00579 а.

Литература

1. Sacread's. URL: <http://sacred.readthedocs.io/en/latest/> (дата обращения: 9.05.2016).
2. Hyperopt. URL: <http://hyperopt.github.io/hyperopt/> (дата обращения: 12.05.2016).
3. FGLab. URL: <https://kaixhin.github.io/FGLab/> (дата обращения: 16.05.2016).
4. Luigi. URL: <http://luigi.readthedocs.io/en/stable/> (дата обращения: 11.05.2016).
5. Kaggle Ensembling Guide. URL: <http://mlwave.com/kaggle-ensembling-guide/> (дата обращения: 11.05.2016).
6. A deep learning approach with an ensemble-based neural

network classifier for black box ICML 2013 Contest. URL: http://deeplearning.net/wp-content/uploads/2013/03/LukaszRomaszko_ICML2013_BlackBox.pdf (дата обращения: 10.05.2016).

7. О проблеме генерации разнообразия ансамблей индивидуальных моделей в задачах идентификации. URL: <http://vsru2014.ipu.ru/proceedings/prcdngs/3214.pdf> (дата обращения: 13.05.2016).

8. Jing Yang, Xiaoqin Zeng, Shuiming Zhong, Shengli Wu. Effective neural network ensemble approach for improving generalization performance. IEEE Transactions on Neural Networks and Learning Systems, 2013, vol. 24, iss. 6, pp. 878–887.

9. Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images. URL: <https://arxiv.org/abs/1603.04833> (дата обращения: 5.05.2016).

10. Learning ensembles of convolutional neural networks. URL: <http://theorycenter.cs.uchicago.edu/REU/2014/final-papers/chen.pdf> (дата обращения: 10.05.2016).

11. Хайкин С. Нейронные сети: полный курс; [пер. с англ.]. М.: Вильямс, 2006. 1104 с.

12. Belyavsky G., Misyura V., Puchkov E. Prediction intervals for time series using neural networks based on wavelet-core. Far East Jour. of Mathematical Sciences, 2016, vol. 100, iss. 3, pp. 413–425.

13. Mohamed Akram Zaytar, Chaker El Amrani. Sequence to sequence weather forecasting with long short-term memory recurrent neural networks. Intern. Jour. of Comp. Applications, 2016, vol. 143, no. 11, pp. 7–11.

14. Li D., Yu Dong, and Platt J. Scalable stacking and learning for building deep architectures. Acoustics, Speech and Signal Processing, 2012.

15. MySQL. URL: <https://www.mysql.com/> (дата обращения: 2.05.2016).

16. InfluxDB. URL: <https://influxdata.com/> (дата обращения: 2.05.2016).

17. NumPy. URL: <http://www.numpy.org> (дата обращения: 3.05.2016).

18. Scikit-learn. URL: <http://scikit-learn.org> (дата обращения: 3.05.2016).

19. CherryPy. URL: <http://www.cherrypy.org> (дата обращения: 7.05.2016).

20. Theano. URL: <http://deeplearning.net/software/theano/> (дата обращения: 17.05.2016).

21. Keras. URL: <https://keras.io/> (дата обращения: 10.05.2016).

22. TensorFlow. URL: <https://www.tensorflow.org/> (дата обращения: 13.05.2016).

23. Пучков Е.В., Белявский Г.И. Разработка нейроэмулятора для решения задач прогнозирования и классификации. Ростов н/Д: Из-во РГСУ, 2012. 138 с.

24. AngularJS. URL: <https://angularjs.org/> (дата обращения: 13.05.2016).

Software & Systems

DOI: 10.15827/0236-235X.117.012-020

Received 11.08.16

2017, vol. 30, no. 1, pp. 12–20

NEURAL NETWORK ENSEMBLES STORAGE DEVELOPMENT

E.V. Puchkov¹, Ph.D (Engineering), Associate Professor, puchkoff@i-intellect.ru

S. Terekhov¹, Graduate Student, isergeiterekhov@gmail.com

¹ DSTU Academy of Civil Engineering and Architecture, Socialisticheskaya St. 162, Rostov-on-Don, 344022, Russian Federation

Abstract. An important tool in the work of a data analysis and machine learning expert is software for an experiment organization. This is primarily related to a large number of stages in data processing and the characteristic aspects of their implementation. In the course of this work the authors have designed and developed a prototype of neural network ensemble storage for data structured storing on various stages of time series forecasting. The article considers a data model, data storage architecture and mechanisms of data acquiring and redistribution in the storage. There is also a description of the developed

class model for software-based interaction with the storage. In order to store data on objects and relationships between these objects there has been used MySQL. For storing time series we used non-relational database InfluxDB. There is also user interface with data visualization and easy interaction with the neural network ensembles storage. The system has been tested using solar activity data in the period from January 1700 to February 2015. The experiment (using LSTM recurrent network) showed that an error of a neural network ensemble was lower than an error of each individual neural network model. LSTM was built using the library Keras, the Blending approach was used to form an ensemble.

The results of this work indicate the prospects of the developed software solution and provide a high degree of integration into scalable Python software. The development of a fully functional system will allow not only organizing the data analysis process, but also improving the performance of resulting models due to ensemble formation process automation.

Keywords: storage, non-relational database, recurrent neural network, LSTM, ensemble, stacking, time series forecasting.

Acknowledgements. The research has been financially supported by RFBR, project no. 14-01-00579 a.

References

1. *Sacred*'s. Available at: <http://sacred.readthedocs.io/en/latest/> (accessed May 9, 2016).
2. *Hyperopt*. Available at: <http://hyperopt.github.io/hyperopt/> (accessed May 12, 2016).
3. *FGLab*. Available at: <https://kaixhin.github.io/FGLab/> (accessed May 16, 2016).
4. *Luigi*. Available at: <http://luigi.readthedocs.io/en/stable/> (accessed May 11, 2016).
5. *Kaggle Ensembling Guide*. Available at: <http://mlwave.com/kaggle-ensembling> (accessed May 11, 2016).
6. *A Deep Learning Approach with an Ensemble-Based Neural Network Classifier for Black Box ICML 2013 Contest*. Available at: http://deeplearning.net/wp-content/uploads/2013/03/LukaszRomaszko_ICML2013_BlackBox.pdf (accessed May 10, 2016).
7. On the problem of individual model assembly variety generation in identification problems. *XII Vseross. soveshchanie po problemam upravleniya VSPU-2014* [All-Russian Conf. on Management Problems VSPU-2014]. Available at: <http://vspu2014.ipu.ru/proceedings/prcdngs/3214.pdf> (accessed May 13, 2016).
8. Yang J., Zeng X., Zhong Sh., Wu Sh. Effective Neural Network Ensemble Approach for Improving Generalization Performance. *IEEE Trans. on Neural Networks and Learning Systems*. 2013, vol. 24, iss. 6, pp. 878–887.
9. *Ensemble of Deep Convolutional Neural Networks for Learning to Detect Retinal Vessels in Fundus Images*. Available at: <https://arxiv.org/abs/1603.04833> (accessed May 5, 2016).
10. *Learning Ensembles of Convolutional Neural Networks*. Available at: <http://theorycenter.cs.uchicago.edu/REU/2014/final-papers/chen.pdf> (accessed May 10, 2016).
11. Haykin S.O. *Neural Networks: A Comprehensive Foundation*. Prentice Hall Publ., 2nd ed., 1998, 842 p. (Russ.ed.: Moscow, Vilyams Publ., 2006, 1104 p.).
12. Belyavsky G., Misyura V., Puchkov E. Prediction intervals for time series using neural networks based on wavelet-core. *Far East Jour. of Mathematical Sciences*. 2016, vol. 100, iss. 3, pp. 413–425.
13. Mohamed Akram Zaytar, Chaker El Amrani Sequence to Sequence Weather Forecasting with Long Short-Term Memory Recurrent Neural Networks. *Int. Jour. of Computer Applications*. 2016, vol. 143, no. 11, pp. 7–11.
14. Li D., Yu D., Platt J. Scalable stacking and learning for building deep architectures. *Acoustics, Speech and Signal Processing*. 2012.
15. *MySQL*. Available at: <https://www.mysql.com/> (accessed May 2, 2016).
16. *InfluxDB*. Available at: <https://influxdata.com/> (accessed May 2, 2016).
17. *NumPy*. Available at: <http://www.numpy.org> (accessed May 3, 2016).
18. *Scikit-learn*. Available at: <http://scikit-learn.org> (accessed May 3, 2016).
19. *CherryPy*. Available at: <http://www.cherrypy.org> (accessed May 7, 2016).
20. *Theano*. Available at: <http://deeplearning.net/software/theano/> (accessed May 17, 2016).
21. *Keras*. Available at: <https://keras.io/> (accessed May 10, 2016).
22. *TensorFlow*. Available at: <https://www.tensorflow.org/> (accessed May 13, 2016).
23. Puchkov E.V., Belyavsky G.I. *Razrabotka neyroemulyatora dlya resheniya zadach prognozirovaniya i klassifikatsii* [Neuroemulator Development to Solve Forecasting and Classification Problems]. Rostov-on-Don, RGSU Publ., 2012, 138 p.
24. *AngularJS*. Available at: <https://angularjs.org/> (accessed May 13, 2016).

Примеры библиографического описания статьи

1. Пучков Е.В., Терехов С. Разработка системы хранения ансамблей нейросетевых моделей // Программные продукты и системы. 2017. Т. 30. № 1. С. 12–20; DOI: 10.15827/0236-235X.117.012-020.

2. Puchkov E.V., Terekhov S. Neural network ensembles storage development. *Programmnye produkty i sistemy* [Software & Systems]. 2017, vol. 30, no. 1, pp. 12–20 (in Russ.); DOI: 10.15827/0236-235X.117.012-020.

УДК 658.26:550.8

Дата подачи статьи: 30.11.16

DOI: 10.15827/0236-235X.117.021-027

2017. Т. 30. № 1. С. 21–27

ИНТЕЛЛЕКТУАЛЬНАЯ ПОДДЕРЖКА ПРИНЯТИЯ РЕШЕНИЙ ПРИ ДИСПЕТЧИРОВАНИИ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ В МНОГОНОМЕНКЛАТУРНОМ МАШИНОСТРОЕНИИ

Г.Б. Бурдо, д.т.н., профессор, зав. кафедрой, gbtms@yandex.ru;

Н.А. Семенов, д.т.н., профессор

*(Тверской государственный технический университет,
наб. Аф. Никитина, 22, г. Тверь, 170026, Россия)*

В последние пятнадцать лет структура машиностроительного и приборостроительного производств претерпела серьезные изменения, обусловленные требованиями заказчиков продукции получать наукоемкие изделия в определенное время. Предприятия указанных отраслей стали разрабатывать и производить одновременно большое число различных изделий, то есть стали многономенклатурными.

Исторически многономенклатурные машиностроительные и приборостроительные предприятия не были оснащены автоматизированными инструментальными средствами, позволяющими эффективно управлять технологическими процессами. Это объясняется высокой динамичностью их производственных систем, отсутствием повторяемости находящихся в изготовлении заказов и возникающих производственных ситуаций, а также влиянием значительного числа случайных факторов, нарушающих нормальный ход технологических процессов. В результате срываются сроки поставки продукции и, таким образом, ухудшаются экономические показатели деятельности предприятий и фирм.

В связи с этим понятна актуальность создания автоматизированных систем поддержки принятия решений в автоматизированных системах управления технологическими процессами.

Диспетчирование технологических процессов имеет своей целью введение их в нормальный график и является одной из важнейших составляющих при управлении ими.

В данной работе реализован комбинированный подход к выработке управляющих воздействий. Исходя из наличия большого числа случайных возмущающих воздействий, в автоматизированной системе выполняется учет наиболее значимых и наиболее вероятных из них.

Поэтому путем сравнения и анализа планируемых и фактических времен (времена начала и окончания) операций технологических процессов, тенденции развития ситуации (накапливание или уменьшение рассогласования) накапливаемым итогом выявляются наиболее вероятные причины невыполнения плана и возможные управляющие воздействия. Анализ производится с помощью базы знаний, построенной на основе производственных моделей. Выявленные причины являются «подсказками» для второго этапа.

На этом этапе с заранее оговоренной периодичностью или при возникновении исключительной ситуации группой экспертов из числа работников предприятия обсуждаются и оцениваются альтернативы. На основании методики нечеткого управления определяется взвешенная оценка уверенности экспертов в достижимости нужного результата реализацией того или иного управляющего воздействия и принимается окончательное решение.

Ключевые слова: машиностроительное многономенклатурное производство, автоматизированная система управления технологическими процессами, производственная модель знаний, нечеткое управление.

Диспетчирование в машиностроении заключается в отслеживании хода выполнения *технологических процессов* (ТП) с целью приведения сроков их окончания в соответствие с требуемыми путем управляющих воздействий [1]. Подавляющее большинство АСУ ТП машиностроительных предприятий *многономенклатурного производства* (МП) различного уровня ориентированы [2, 3] лишь на фиксацию хода выполнения дискретных ТП механической обработки (их *календарных планов-графиков* (КПГ)) за период *интервала оперативного планирования* (ИОП). При этом их основная функция – выработка управляющего воздействия на производственную систему – остается за человеком, то есть за диспетчером [4, 5]. В МП проблема оперативного управления усугубляется большой номенклатурой заказов, находящихся в изготовлении в технологических подразделениях, а также большим числом объективных и субъективных случайных факторов, влияющих на фактические

сроки выполнения заказов. Поэтому целью исследования явилось создание интеллектуальных механизмов поддержки принятия решений, позволяющих формировать рекомендации – подсказки для управленческого персонала в МП при диспетчировании ТП [6] с целью выработки согласованного управляющего воздействия.

Модель диспетчирования ТП

Имеется технологическое подразделение механической обработки определенной структуры, для которого сформированы объемный и календарный планы, и КПГ по выпуску изделий. Для данного подразделения необходимо диспетчированием обеспечить выполнение КПГ, определяющего выполнение и остальных планов. Экспертным опросом установлено, что для оценки ситуации в подразделении объектами диспетчирования в МП должны являться не только сроки запуска и вы-

пуска изделий, но и сроки выполнения отдельных технологических операций для тех деталей, цикл изготовления которых находится в нескольких ИОП. Второй группой контролируемых параметров является фактическая численность работающих и оборудования с разбивкой по типам и группам. Также с помощью экспертов были определены возможные причины отклонений от КПП и возможные корректирующие управляющие воздействия.

На шаге 1 работы АСУ ТП выполняются этапы 1 и 2.

1. На первом этапе определяются множества фактических сроков поступления партий деталей на первую операцию в пределах данного ИОП – $\{T1_{lqi}\}$ (l – номер партии деталей, q – порядковый номер операции, i – группа оборудования). Определяется множество интервалов $\Delta P1_{lqi}$ расхождения сроков с плановыми сроками поступления $T2_{lqi}$: $T1_{lqi} - T2_{lqi} = \Delta P1_{lqi}$, и сравнивается с допустимым $[\Delta P1]$ значением (при задержке сроков): $\Delta P1_{lqi} \leq [\Delta P1]$.

Рекомендуется $[\Delta P1]$ принимать равным до 50 % от длительности смены. Далее рассчитывается среднее и объемное отставания (опережения) по технологическому подразделению:

$$\Delta P2 = \sum_1^d (\Delta P1_{lqi}) / d ; \Delta P3 = \sum_1^d (\Delta P1_{lqi}) ,$$

где d – число запущенных партий деталей.

Величина $\Delta P2$ сравнивается с допустимым интервалом: $\Delta P2 \leq [\Delta P2]$.

2. На втором этапе производится контроль выполнения КПП за каждую рабочую смену. Определяется множество фактических времен окончания последней операции для партий деталей $\{T3_{lqi}\}$ и сравнивается с плановыми $\{T4_{lqi}\}$: $T3_{lqi} - T4_{lqi} = \Delta P4_{lqi}$, где $\Delta P4_{lqi}$ – расхождение времени окончания обработки l -й партии деталей по q -й операции на i -й группе станков относительно плановой. По результатам отклонений сроков завершения работ рассчитывается среднее и объемное отставание (опережение) по технологическому подразделению:

$$\Delta P5 = \sum_1^d (\Delta P4_{lqi}) / d ; \Delta P6 = \sum_1^d (\Delta P4_{lqi}) .$$

В случае отставания от плана величины $\Delta P4_{lqi}$ и $\Delta P5$ сравниваются с допустимой величиной интервала: $\Delta P5 \leq [\Delta P5]$, $\Delta P4_{lqi} \leq [\Delta P4]$. Так как $\Delta P6$ объединяет в себе все виды работ, соответствующие типам и группам станков, производят дифференциацию на подмножества $\{\Delta P6_j\}$ и $\{\Delta P6_{ji}\}$:

$$\sum_1^m \Delta P6_j = \Delta P6 , \sum_1^n \Delta P6_{ji} = \Delta P6_{ji} ,$$

где j – номер типа; i – номер группы.

Определяется расчетный (по плановому числу рабочих и рабочих мест) фонд времени работы оборудования $\Phi 1$ по технологическому подразделению по типам станков $\{\Phi 1_j\}$ и группам $\{\Phi 1_{ji}\}$.

Аналогичные показатели рассчитываются и исходя из фактической численности рабочих и рабо-

чих мест: $\Phi 2$, $\{\Phi 2_j\}$, $\{\Phi 2_{ji}\}$. Фонды времени рассчитываются накапливающимся итогом за период времени от начала ИОП до конца контролируемой смены (1 смена – t_1 , 2 – t_2 , ..., k – t_k , ..., c – t_c ; c – число смен в ИОП).

Отнесение объемных отклонений от КПП к соответствующим фондам времени будет более эффективно характеризовать выполнение КПП и тенденцию изменения отклонений. Введем обозначения:

$$\alpha 1(t_k) = \Delta P6(t_k) / \Phi 1 ; \alpha 2(t_k) = \Delta P6(t_k) / \Phi 2(t_k) ;$$

$$\alpha 3_j(t_k) = \Delta P6_j(t_k) / \Phi 1_j ;$$

$$\alpha 4_j(t_k) = \Delta P6_j(t_k) / \Phi 2_j(t_k) ;$$

$$\alpha 5_{ji}(t_k) = \Delta P6_{ji}(t_k) / \Phi_{ji}^p ;$$

$$\alpha 6_{ji}(t_k) = \Delta P6_{ji}(t_k) / \Phi 2_{ji}(t_k) ,$$

где величина в скобках (t_k) означает, что параметры определяются на соответствующие моменты времени (t_1, t_2, \dots, t_c).

Указанные параметры, как и средние и объемные отставания (опережения) от плана по технологическому подразделению, группам и типам рабочих мест, используются для выработки возможных управляющих воздействий на шаге 2.

Модель разработки возможных управляющих воздействий

Анализ полученного на шаге 1 множества информационных параметров выполняется после завершения каждой смены на шаге 2 работы АСУ ТП. Возможные управляющие воздействия формируются на основе сравнения результатов выполнения КПП на момент окончания k -й смены (время t_k) с результатами по окончании предыдущей ($t-1$)-й смены (t_{k-1}) и с состоянием t_0 (момент запуска) с использованием 15 установленных правил (ПР) – базой знаний [7], отражающих распознанную ситуацию в технологическом подразделении и не противоречащих друг другу [8]. Впоследствии эксперты обсуждают возможные управляющие воздействия с целью определения наиболее рационального.

ПР1. ЕСЛИ (выполняются условия:

У1) $\Delta P5(t_k) \leq [\Delta P5]$ И

У2) $\forall i, \Delta P4_{lqi}(t_k) \leq [\Delta P4]$ И

У3) $\Delta P5(t_k) \geq 0$), (1)

ТО (отклонения от графика при положительном $\Delta P5(t_k)$ находятся в допустимых пределах И технологические подразделения продолжают функционировать в прежнем режиме). Условие (1) определяет нормальное функционирование технологического подразделения.

ПР2. ЕСЛИ (условие У3) в ПР1 несправедливо (то есть $\Delta P5(t_k) \geq 0$, идет опережение КПП и имеются резервы дозагрузок работами, имеющимися в распоряжении мастеров), ТО (выявляются группы станков $\{i\}$, для которых $\Delta P4_{lqi}(t_k) < 0$, И определя-

ется резерв фонда времени по каждой указанной группе (РГ_г) И суммарный: $\{РГ_i\} = \{\sum_1^d (\Delta P1_{lqi})\}$.

Резерв определяет величины станкочасов работ, которыми можно догрузить группы оборудования технологического подразделения.

ПР3. ЕСЛИ ($\{РГ_i\} > 0$ И есть локальные резервы), ТО догрузить группы станков с $РГ_i > 0$.

ПР4. ЕСЛИ ($\{РГ_i\} > 0$ И нет локальных резервов у мастеров), ТО передать работы из (q+1)-го ИОП.

ПР5. ЕСЛИ $\Delta P5(t_k) > [\Delta P5]$, ТО имеются отставания от планового КПП. Далее выясняется момент появления отставания путем сравнения состояния с предыдущим $k = k - 1$ по объемным и средним отклонениям.

ПР6. ЕСЛИ (выполняются

$$У1) \Delta P5(t_k) - \Delta P5(t_{k-1}) \leq [\Delta P5] \text{ И}$$

$$У2) \Delta P6(t_k) / \Delta P6(t_{k-1}) = 0,9, \dots, 1,1), \quad (2)$$

ТО накопление отставания за k-ю смену не происходило.

Далее сравнивается текущее состояние с состоянием t_0 запуска.

ПР7. ЕСЛИ (выполняются

$$У1) \Delta P5(t_k) - \Delta P2(t_0) \leq [\Delta P5] \text{ И}$$

$$У2) \Delta P6(t_k) / \Delta P6(t_0) = 0,9, \dots, 1,1), \quad (3)$$

ТО (отставание получилось за счет невыполнения сроков запуска И необходимо число дополнительных смен, $\{C_{ли}\}$).

Определяются объемы множества отставаний работ по каждому типу станков j – необходимый дополнительный фонд работы $\Phi_{dj} = \sum_i \Delta P4_{lqi}$, делением на 8 получают необходимое число дополнительных смен по каждому типу оборудования для сокращения отставания: $C_{ли} = \Phi_{dj} / 8$.

ПР8. ЕСЛИ (в (2) И (3) не выполняется условие У1) или У2), ТО переход в «ручной» режим (некорректная ситуация, вызванная вводом неверных данных).

ЕСЛИ (ПР7 неверна) (то есть состояние t_k отлично от предшествующего), ТО далее анализируются причины.

ПР9. ЕСЛИ (состояние t_k отлично от предшествующего t_{k-1}) И $\alpha 1(t_k) \leq \alpha 2(t_k)$ (то есть фактический фонд времени работы был больше), ТО (множество оцениваемых причин: неверная оценка времен, неудовлетворительная работа мастеров, простой по организационным причинам).

ПР10. ЕСЛИ (состояние t_k отлично от предшествующего t_{k-1} И выполняется:

$$\alpha 1(t_k) > \alpha 2(t_k), \quad (4)$$

ТО (ЕСЛИ

$$0,9 \times \Delta P5(t_k) \leq \Phi 1(t_k) \geq \Phi 2(t_k) \times K3)) \quad (5)$$

(оцениваются возможные отставания за счет уменьшения фонда времени работы: где K_3 – средний коэффициент загрузки) – **ПР11**), ИНАЧЕ **ПР12**.

ПР11. ЕСЛИ условие (5) выполняется, ТО (отставание от КПП произошло из-за отличия фондов времени), ИНАЧЕ из-за отличия фондов времени и организационных потерь. Рассчитывается коэффициент веса оргпотерь (K_o) по отношению к общим потерям:

$$K_o = (\Delta P_T^K(t_T) - K_3 \times (\Phi_o^P(t_T) - \Phi_o^P(t_i))) / (K_3 \times (\Phi_o^P(t_T) - \Phi_o^P(t_i))).$$

ПР12. ЕСЛИ выполняется (4), ТО (анализируются фонды времени работы оборудования по типам и группам). Попарно сравниваются коэффициенты $\alpha 3_j$ и $\alpha 4_{Tj}$:

$$\forall j, (j=1, n), (\alpha 4_j(t_k) / \alpha 3_j(t_k)) \geq 1,05, \quad (6)$$

где 1,05 определяет возможное повышение выработки за счет реализации локальных резервов и перевыполнения норм. Такие типы станков запоминаются (индексы $\{j'\} \in \{j\}$). Выявленные j'-е типы далее проверяются по соотношению $\alpha 6_{ji}$ и $\alpha 5_{ji}$ для i-х групп:

$$\forall j_i, i = \overline{1, m}, (\alpha 6_{ji}(t_k) / \alpha 5_{ji}(t_k)) > 1,05). \quad (7)$$

Отвечающие условию $\{j'\} \in \{j\}$ запоминаются. ЕСЛИ (выявленные множества $\{j'\}$ и $\{i'\}$ определяют более 80 % отставания от КПП), то есть

$$\sum_j \sum_{i'} (\Delta P6_{ji'}) \geq 0,8 \times \Delta P6, \quad (8)$$

ТО анализу подлежат только они.

0,8 взято исходя из плановой загрузки станков, возможности устранения остальных отставаний за счет локальных резервов. Далее определяются отличия фондов времени по группам и по типам:

$$\{\Delta \Phi_{ji'}(t_k)\} = \{\Phi 1_{ji'}(t_k) - \Phi 2_{ji'}(t_k)\},$$

$$\{\Delta \Phi_j(t_k)\} = \{\Phi 1_j(t_k) - \Phi 2_j(t_k)\} = \sum_i \Phi_{ji'}(t_k).$$

По каждой группе и типу станков находятся $\beta_{ji'}(t_k)$ и $\beta_j(t_k)$, определяющие соотношение фондов времени:

$$\beta_{ji'}(t_k) = \Delta \Phi_{ji'}(t_k) / \Phi 2_{ji'}(t_k);$$

$$\beta_j(t_k) = \Delta \Phi_j(t_k) / \Phi 1_j(t_k). \quad (9)$$

ПР13. ЕСЛИ $\beta_{ji'}(t_k) < 0,1$ (0,1 учитывает возможности производства по ликвидации отставаний от КПП), ТО (необходимы мобилизация локальных резервов, допущение переработок в пределах $\Delta \Phi_{ji'}$ И перераспределение работ между взаимозаменяемыми типами или группами станков с учетом $\{\beta_{ji'}\}$).

ПР14. ЕСЛИ $\beta_j(t_k) \geq 0,1$, ТО (возможно принятие решений из следующих: 1) работа в выходные дни, следующие за ИОП, по группам оборудования, 2) увеличение фактического фонда времени работы оборудования путем отзыва из отпусков рабочих, приема на работу и т.д., 3) приобретение оборудования и прием рабочих при отсутствии свободного оборудования, то есть создание дополнительных рабочих мест по типам ΔPM_j и группам

ΔPM_{ji} станков; 4) передача части работ другим технологическим подразделениям или сторонним организациям).

1-е решение возможно, если

$$(\beta_{j'i'}(t_k) > 0,1) \wedge ((\Delta\Phi_{j'i'}(t_k)) / 8) \leq 2) \quad (2 \text{ смены}).$$

Дополнительное число рабочих по типам оборудования, $\Delta C_j(t_k)$: $\{\Delta C_j(t_k)\} = \{\Phi_{j'}(t_k) / (8 \times m)\}$, (m – индекс дня ИОП (1, 2 и т.д.)), и группам: $\{\Delta C_{ji}(t_k)\} = \{\Delta\Phi_{j'i'}(t_k) / (8 \times m)\}$;

$$\sum (\Delta C_{ji}(t_i) = \Delta C_j(t_i)).$$

2-е решение возможно, если

$$(\beta_{j'i'}(t_k) > 0,1) \wedge ((\Delta\Phi_{j'i'}(t_i) / 8) > 2).$$

Прерогатива принятия **решений 3 и 4** (объемы передаваемых работ пропорциональны величинам $\Delta\Phi_{j'i'}(t_k)$) – высшее управленческое звено фирмы. Решение 1 может быть принято на основе анализа 1–2 ИОП, решение 2 требует анализа 3–5 ИОП, для решений 3 и 4 необходим анализ в течение 1–2 месяцев.

ПР 15. ЕСЛИ (неверны (6) ИЛИ ((7) И (8)), ТО (переход к анализу различия фондов времени по типам и группам оборудования:

$$\{\Delta\Phi_{ji}(t_k)\} = \{\Phi 1_{ji}(t_k) - \Phi 2_{ji}(t_k)\};$$

$$\{\Delta\Phi_j(t_k)\} = \{\Phi 1_j(t_k) - \Phi 2_j(t_k)\}$$

И определяют коэффициенты $\beta_{ji}(t_k)$ и $\beta_j(t_k)$ по (9)). Дальнейший анализ ситуации выполняется аналогично. Заметим, что все приведенные числовые значения справедливы для условий базового предприятия, уточняются на основе опыта работы системы и для других предприятий могут быть рекомендованы как начальные. Некоторые отличия имеются на последнем интервале t_k , соответствующем окончанию ИОП. Здесь необходимо множеству отставаний от КПП на данном ИОП $\{\Delta P_6\}$ присвоить приоритеты [9], сформировать множества невыполненных работ для включения их в следующий ИОП.

Процедуры оценки возможных управленческих воздействий

Полученные с помощью производственной базы знаний причины отклонения хода выполнения ТП от расчетного КПП и возможные управляющие воздействия анализируются «внутренними» экспертами на шаге 2 работы АСУ ТП. Анализ осуществляется после окончания каждого ИОП.

Процедуры выполняются работниками организации (внутренние эксперты) при активном управлении ЛПР. Как показал опыт, наиболее качественный состав экспертов – это работники самой организации, по своим должностным обязанностям относимые к управленческому персоналу (с точки зрения производства) различного уровня. Состав внутренних экспертов, как правило, следующий:

субъект 1 – ЛПР – руководитель организации по производству или его заместитель;

субъект 2 – начальник производственной системы;

субъект 3 – руководитель (руководители) подразделения (подразделений);

субъект 4 – руководитель (руководители) диспетчерской службы (служб).

Алгоритм принятия решений коллективом внутренних экспертов состоит в следующем.

1. Внутренним экспертам присваиваются веса (их сумма равна 1) с учетом степени ответственности за принимаемое решение и квалификации в предметной области решения.

2. Определяется допустимая нижняя граница общей уверенности внутренних экспертов в эффективности управляющего воздействия (на основе опыта работы рекомендуется $\geq 0,7$).

3. Определяется периодичность корректировок (рекомендуется по окончании каждого ИОП).

4. Определяются точки оценки возможных решений и управляющих воздействий (рекомендуется после шага 1 и шага 2 работы системы); $i=1$.

5. Переход к i -му периоду (ИОП) корректировок.

6. Обсуждение и оценка варианта управляющего воздействия (уверенность в достижении результата реализацией управляющего воздействия) каждым внутренним экспертом в числовой форме (от 0 до 1).

7. Получение взвешенной оценки общей уверенности внутренних экспертов в достижении результата реализацией управляющего воздействия.

8. Принятие решения ЛПР и переход к следующей точке оценки.

9. $i = i + 1$. Возврат на пункт 5.

Отдельного внимания заслуживает способ определения взвешенной уверенности всех внутренних экспертов. С этой целью исходя из экспертных оценок была разработана база знаний на основе нечетких множеств [9].

Правила имеют вид:

(ЕСЛИ уверенность эксперта 1 с весом P_1 равна A_1) И (ЕСЛИ уверенность эксперта 2 с весом P_2 равна A_2) И (ЕСЛИ уверенность эксперта 3 с весом P_3 равна A_3) И

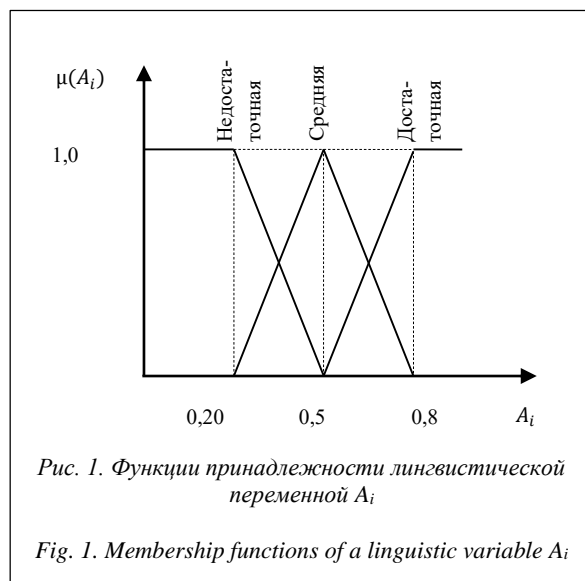
...

(ЕСЛИ уверенность эксперта n с весом P_n равна A_n), ТО (взвешенная уверенность экспертов в результате равна B_i).

Лингвистическая переменная A_i имеет 3 терма: достаточная, средняя, недостаточная (рис. 1).

Лингвистическая переменная B_i имеет 5 термов: весьма надежная, надежная, сомнительная, ненадежная, весьма ненадежная (рис. 2).

Рассмотрим, как строится производственная база нечетких правил определения взвешенной уверенности экспертов в эффективности выбранного управляющего воздействия.



С этой целью были введены следующие правила создания моделей.

1. Веса внутренних экспертов P_i принимаются из ряда чисел, кратных k_i раз минимальному из них (k_i – целое число, равное или большее 1).

$$P_{\min} = \min\{P_1, P_2, \dots, P_i, \dots, P_n\}.$$

2. Делается предварительное преобразование подусловий по введенному правилу, когда одно подусловие заменяется k_i подусловиями, соединенными оператором «И».

Например, исходное подусловие имеет вид:

(ЕСЛИ уверенность эксперта i с весом ($k_i \times P_{\min}$) равна A_i) И...

Заменяющие подусловия:

(ЕСЛИ уверенность эксперта i с весом P_{\min} равна A_i И

ЕСЛИ уверенность эксперта i с весом P_{\min} равна A_i И

... ЕСЛИ уверенность эксперта i с весом P_{\min} равна A_i) И...

То есть исходное подусловие разбивается на k_i подусловий.

3. Записываются с учетом п. 2 нечеткие продукционные правила определения взвешенной уверенности экспертов.

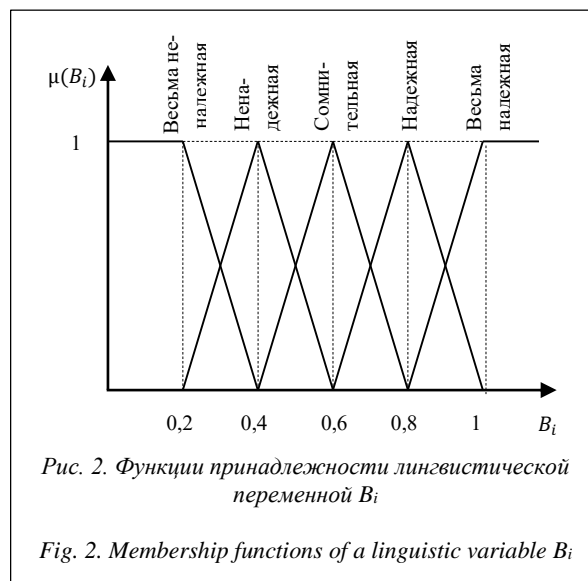
4. В продукционном правиле взаимоисключаются подусловия, содержащие значения лингвистической переменной A_i (достаточная \Leftrightarrow недостаточная).

Например, исходное правило:

(ЕСЛИ уверенность эксперта 1 – достаточная И уверенность эксперта 1 – недостаточная И уверенность эксперта 2 – недостаточная И уверенность эксперта 3 – достаточная И уверенность эксперта 3 – достаточная И уверенность эксперта 4 – средняя), ТО...

Преобразованное правило:

ЕСЛИ уверенность эксперта 1 – достаточная И уверенность эксперта 3 – достаточная И уверенность эксперта 3 – достаточная И уверенность эксперта 4 – средняя), ТО...



Как видим, в результате получаются продукционные правила, содержащие только смежные значения лингвистической переменной A_i .

С помощью экспертных оценок сформированы следующие нечеткие продукционные правила, определяющие значение лингвистической переменной B_i (взвешенная уверенность экспертов).

1. ЕСЛИ все подусловия содержат значение лингвистической переменной A_i – достаточная, ТО значение лингвистической переменной B_i – весьма надежная.

2. ЕСЛИ все подусловия содержат значение лингвистической переменной A_i – недостаточная, ТО значение лингвистической переменной B_i – весьма ненадежная.

3. ЕСЛИ число подусловий со значением лингвистической переменной A_i – средняя (превышает число подусловий со значением лингвистической переменной A_i – достаточная или недостаточная) ИЛИ (не более, чем на одно меньше числа подусловий со значением лингвистической переменной A_i – достаточная или недостаточная), ТО лингвистическая переменная B_i – сомнительная.

4. ЕСЛИ (число подусловий со значением лингвистической переменной A_i – достаточная (или недостаточная) более, чем на одно, превышает число подусловий со значением лингвистической переменной A_i – средняя), ТО лингвистическая переменная B_i – надежная (или ненадежная) соответственно.

Например: (ЕСЛИ уверенность эксперта 1 – достаточная И уверенность эксперта 3 – достаточная И уверенность эксперта 3 – достаточная И уверенность эксперта 4 – средняя), ТО значение лингвистической переменной B_i – надежная.

Fuzzy-логическая конъюнкция подусловий выполняется по правилу минимума.

Нахождение функций совместной принадлежности – по правилу нечеткой импликации Мамдани.

Дефазификация взвешенной оценки уверенности экспертов (результат) выполняется по центральному методу.

Заключение

Таким образом, на основе автоматизированного распознавания производственной ситуации по формальным признакам системой генерируются возможные решения, оцениваемые впоследствии экспертами. Это позволило уточнить выбор управляющего воздействия. При оценке управляющего воздействия эксперты учитывают дополнительные (по отношению к системе) параметры, так как в условиях производственных систем многоменклатурного машиностроения в моделях АСУ ТП учет всех возмущающих действий практически невозможен.

Разработанные методики и модели диспетчирования в рамках АСУ ТП [10, 11] успешно прошли промышленную апробацию на базовом предприятии и могут быть рекомендованы к реализации модулем в автоматизированных промышленных системах планирования и управления производственными подразделениями классов ERP, MRP-II и Scada [3].

Следует подчеркнуть положительный эффект и от привлечения цехового персонала к процессу принятия решений, что увеличивает их меру ответственности за результаты работы и заставляет глубже вникать в суть проблем управления производством.

Опыт реализации методик показал их доступность для понимания и использования цеховым управленческим персоналом.

Работа выполнена при финансовой поддержке РФФИ, проект № 14-01-00324.

Литература

1. Самойлович В.Г. Организация производства и менеджмент. М.: Академия, 2008. 336 с.
2. Балахонова И.В., Волчков С.А., Капитуров В.А. Интеграция процессов с помощью ERP-системы. Н. Новгород: Приоритет, 2006. 464 с.
3. Обухов И.А., Гайфуллин Б.Н. Автоматизация систем управления предприятиями стандарта ERP-MRP-11. М.: Интерфейс-Пресс, 2002. 286 с.
4. Гаврилов Д.А. Календарное планирование: что и как? // Умное производство. 2009. № 4. С. 49–57.
5. Лири Д.О. ERP-системы. Современное планирование и управление ресурсами предприятия. М.: Вершина, 2004. 272 с.
6. Бурдо Г.Б., Семенов Н.А., Исаев А.А. Автоматизированная система управления технологическими процессами в многоменклатурных производствах // Программные продукты и системы. 2012. № 1. С. 80–83.
7. Рыбина Г.В. Основы построения интеллектуальных систем. М.: Финансы и статистика; Инфра-М, 2010. 432 с.
8. Ларичев О.И., Мечитов А.И., Мошквич Е.М., Фуремс Е.М. Выявление экспертных знаний; [под ред. С.В. Емельянова]. М.: Наука, 1989. 128 с.
9. Заде Л. Понятие лингвистической переменной и ее применение к понятию приближенных решений. М.: Мир, 1976. 167 с.
10. Бурдо Г.Б., Палюх Б.В., Семенов Н.А. Системы управления технологическими процессами в машиностроении: подходы к созданию интеллектуальных систем управления: монография. Palmarium Acad. Publ., 2012. 187 с.
11. Бурдо Г.Б. Интеллектуальные процедуры планирования и управления в производственных системах геофизического приборостроения // Программные продукты и системы. 2011. № 3. С. 107–110.

Software & Systems

DOI: 10.15827/0236-235X.117.021-027

Received 30.11.16

2017, vol. 30, no. 1, pp. 21–27

INTELLIGENT DECISION SUPPORT IN PROCESS SCHEDULING IN DIVERSIFIED ENGINEERING

G.B. Burdo¹, Dr.Sc. (Engineering), Head of Chair, gbtms@yandex.ru

N.A. Semenov¹, Dr.Sc. (Engineering), Professor

¹ Tver State Technical University, Nikitin Quay 22, Tver, 170026, Russian Federation

Abstract. In the last fifteen years the structure of machine-building and instrument-making production has undergone major changes due to the requirements of customers to receive high-tech products at a certain time. This fact made relevant companies to design and manufacture a large number of different products simultaneously. It has led them to diversification.

Historically, diversified engineering and instrumentation enterprises were not equipped with automated tools to manage technological processes effectively. This fact might be explained by high acceleration capacity of their production systems, lack of repeatability in a production list and manufacturing situations, as well as influence of random factors that violate a normal process flow status. All this leads to elongation and disruption of product delivery time, and as a result, to the deterioration of financial and economic performance data of enterprises and firms.

In this regard, it becomes clear that creation of automated decision-making support systems in automated technological process control systems is an important problem.

Dispatching of technological process is focused on their introduction into a normal schedule. It is one of the most important components in management.

In this work we implemented a combined approach to making controlling actions. Based on a large number of random disturbances, an automated system records the most important and most probable of them.

Therefore, by comparing and analyzing planned and actual times (start and end times) of technological process operations, possible situation development (accumulation or reduction of disagreement) the system accumulates the results and identifies the most likely causes of plan failure and possible control actions. The analysis is performed using a knowledge base constructed on the basis of production models. The identified causes are “tips” for the second phase.

At this stage with a predetermined frequency or at the occurrence of the exception a group of experts from company employees discusses and evaluates alternatives. Fuzzy control defines a weighted assessment of experts’ confidence in achievability of a desired result by executing various control action and the final decision is accepted.

Keywords: multiproduct machine-building manufacture, the automated control system of technological processes, production knowledge model, fuzzy control.

Acknowledgements. The work has been financially supported by RFBR, project no. 14-01-00324.

References

1. Samoylovich V.G. *Organizatsiya proizvodstva i menedzhment* [Production Organization and Management]. Moscow, Akademiya Publ., 2008, 336 p.
2. Balakhonova I.V., Volchikov S.A., Kapituov V.A. *Integratsiya protsessov s pomoshchyu ERP-sistemy* [Integration of processes using ERP system]. N. Novgorod, Prioritet Publ., 2006, 464 p.
3. Obukhov I.A., Gayfullin B.N. *Avtomatizatsiya sistem upravleniya predpriyatiyami standart ERP-MRP-11* [Automation of business management systems of standard ERP-MRP-11]. Moscow, Interfeys-Press, 2002, 286 p.
4. Gavrilov D.A. Calendar planning: what and how? *Umnoe proizvodstvo* [Intelligent Manufacturing]. 2009, no. 4, pp. 49–57 (in Russ.).
5. Liri D.O. *ERP-sistemy. Sovremennoe planirovanie i upravlenie resursami predpriyatiya* [ERP-systems. Modern planning and resource management in a company]. Vershina Publ., 2004, 272 p.
6. Burdo G.B., Semenov N.A., Isaev A.A. Systems technological processes control in multiproduct manufactures. *Programmnye produkty i sistemy* [Software & Systems]. 2012, no. 1 (97), pp. 80–83 (in Russ.).
7. Rybina G.V. *Osnovy postroyeniya intellektualnykh sistem* [Basics of Intelligent System Constructing]. Moscow, Finansy i statistika. Infra-M Publ., 2010, 432 p.
8. Larichev O.I., Mechitov A.I., Moshkovich E.M., Furems E.M. *Vyyavlenie ekspertnykh znany* [Expert Knowledge Detection]. S.V. Emelyanov (Ed.), Moscow, Nauka Publ., 1989, 128 p.
9. Zadeh L. *Ponyatie lingvisticheskoy peremennoy i eyo primeneniye k ponyatiyu priblizhennykh resheniy* [The concept of a Linguistic Variable and its Application to Creating Approximate Solutions]. Moscow, Mir Publ., 1976.
10. Burdo G.B., Palyukh B.V., Semenov N.A. *Sistemy upravleniya tekhnologicheskimi protsessami v mashinostroyeni: podkhody k sozdaniyu intellektualnykh sistem upravleniya* [Mechanical Engineering Production Process Control Systems: approaches to Intelligent Control System Creation]. Monograph. Palmarium Academic Publ., Saarbruchen, Germany, 2012, 187 p.
11. Burdo G.B. Intelligent procedures of planning and control in geophysical industry production systems. *Programmnye produkty i sistemy* [Software & Systems]. 2011, no. 3 (95), pp. 107–110 (in Russ.).

Примеры библиографического описания статьи

1. Бурдо Г.Б., Семенов Н.А. Интеллектуальная поддержка принятия решений при диспетчировании технологических процессов в многономенклатурном машиностроении // Программные продукты и системы. 2017. Т. 30. № 1. С. 21–27; DOI: 10.15827/0236-235X.117.021-027.
2. Burdo G.B., Semenov N.A. Intelligent decision support in process scheduling in diversified engineering. *Programmnye produkty i sistemy* [Software & Systems]. 2017, vol. 30, no. 1, pp. 21–27 (in Russ.); DOI: 10.15827/0236-235X.117.021-027.

УДК 007:519.816

DOI: 10.15827/0236-235X.117.028-033

Дата подачи статьи: 17.11.16

2017. Т. 30. № 1. С. 28–33

РЕАЛИЗАЦИЯ МЕТОДОВ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ НА ОСНОВЕ ТЕМПОРАЛЬНЫХ РАЗЛИЧИЙ И МУЛЬТИАГЕНТНОГО ПОДХОДА ДЛЯ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ РЕАЛЬНОГО ВРЕМЕНИ

А.П. Еремеев, д.т.н., профессор, eremeev@appmat.ru;

А.А. Кожухов, аспирант, saanchezzzz@yandex.ru

*(Национальный исследовательский университет «Московский энергетический институт»,
ул. Красноказарменная, 14, г. Москва, 111250, Россия)*

В работе описана реализация методов обучения с подкреплением на основе временных (темпоральных) различий и мультиагентной технологии. Рассмотрены возможности комбинирования методов обучения со статистическими и экспертными методами прогнозирования с целью последующей интеграции в инструментальную программную среду для использования в современных перспективных интеллектуальных системах реального времени типа интеллектуальных систем поддержки принятия решений реального времени.

Даны анализ методов обучения с подкреплением (RL-обучения) в плане использования в интеллектуальных системах реального времени, их основные компоненты, преимущества и решаемые задачи. Основное внимание уделено методам RL-обучения на основе временных (темпоральных) различий (TD-методам), разработаны соответствующие алгоритмы. Рассмотрены возможности включения методов RL-обучения в мультиагентную среду и их комбинирования со статистическими и экспертными методами прогнозирования с целью последующей интеграции в инструментальную среду для использования в интеллектуальных системах реального времени типа интеллектуальных систем поддержки принятия решений реального времени для управления и диагностики сложных технических объектов.

Разработана архитектура прототипа подсистемы прогнозирования, включающая эмулятор, моделирующий состояние проблемной области (объекта и внешнего окружения), и модули прогнозирования, анализа и принятия решений, RL-обучения. Выполнена программная реализация прототипа подсистемы прогнозирования с применением мультиагентного подхода для решения задачи экспертного диагностирования сложного технического объекта.

Результаты тестирования и апробации разработанной системы показали ее достаточную эффективность и целесообразность включения в состав современных интеллектуальных систем поддержки принятия решений реального времени.

Ключевые слова: искусственный интеллект, интеллектуальная система, реальное время, обучение с подкреплением, прогнозирование, поддержка принятия решений, программное средство.

Методы обучения с подкреплением (reinforcement learning, RL) [1], основанные на использовании большого количества информации для обучения в произвольной окружающей среде, являются одной из наиболее активно развиваемых областей искусственного интеллекта, связанных с разработкой перспективных *интеллектуальных систем реального времени* (ИС РВ), типичными примерами которых являются *интеллектуальные системы поддержки принятия решений реального времени* (ИСППР РВ) [2, 3].

Одним из наиболее перспективных в плане использования в ИС РВ, относящихся к классу динамических интеллектуальных систем [4, 5], является обучение на основе темпоральных различий (temporal-difference, TD) [1], когда процесс обучения основывается непосредственно на получаемом опыте без предварительных знаний о модели поведения окружающей среды. Ключевой особенностью TD-алгоритмов является обучение на основе различий во временных последовательных предсказаниях. TD-методы, предназначенные для многомерных временных рядов, способны обновлять расчетные оценки, основанные в том числе и на других полученных оценках, не дожидаясь окончательного результата, то есть являются самонастра-

иваемыми. Последнее свойство весьма важно для ИС семиотического типа, способных адаптироваться (подстраиваться) к изменениям в управляемом объекте и/или окружающей среде [3].

Использование мультиагентного подхода в динамических ИС, в том числе ИС РВ (ИСППР РВ), системах распределенного управления и системах интеллектуального анализа данных, способного улучшить эффективность и надежность таких систем, является быстроразвивающимся и перспективным подходом [6].

При разработке современных ИС РВ большое внимание должно быть уделено также средствам прогнозирования развития ситуации на объекте и последствий принимаемых решений, экспертным методам и средствам обучения [3, 7]. Эти средства необходимы для модификации и адаптации ИС РВ к изменениям на объекте и во внешней среде, а также для расширения области применения и улучшения эффективности функционирования систем.

Далее будет дан анализ ряда методов обучения с подкреплением, в частности TD-методов, в плане их последующей интеграции в инструментальную среду для ИС РВ типа ИСППР РВ с применением мультиагентного подхода.

Методы обучения с подкреплением

Будем предполагать, что неопределенность поступающей в БД ИС РВ информации о текущем состоянии проблемной области, объекта и окружающей среды связана в основном с ошибочной работой датчиков (сенсоров) или ошибками соответствующего оперативно-диспетчерского персонала (лиц, принимающих решения, ЛПР). В функции RL-обучения входит адаптация немарковской модели принятия решений к сложившейся ситуации за счет анализа предыстории процесса принятия решений, вследствие чего повышается качество принимаемых решений [1, 8, 9].

В RL-обучении модуль принятия решений, способный посредством взаимодействия с внешней средой и анализа оценочной функции (функции платежа) корректировать стратегию принятия решений, называется *агентом*. Задачей агента является нахождение в процессе обучения оптимальной (для марковского процесса) или приемлемой (если процесс не является марковским) стратегии принятия решений, называемой также *политикой*. Интеллектуальный агент должен уметь поддерживать несколько путей обучения и адаптировать накопленный опыт к изменениям в окружающей среде. В RL-обучении взаимодействие «агент–окружающая среда» моделируется посредством контроллера, связывающего ИС и среду. Процесс восприятия отображает состояния среды (проблемной области) во внутренние представления агента, а процесс воздействия отображает предлагаемые агентом воздействия в действия (преобразования) внешней среды. Обобщенная схема взаимодействия «агент–окружающая среда» приведена на рисунке 1.

Целью RL-обучения является максимизация ожидаемой выгоды R_t , которая определяется как некоторая функция, заданная на последовательности вознаграждений: $R_t = r_{t+1} + r_{t+1} + \dots + r_{t+T}$, где T – завершающий временной шаг; r_t – вознаграждение на

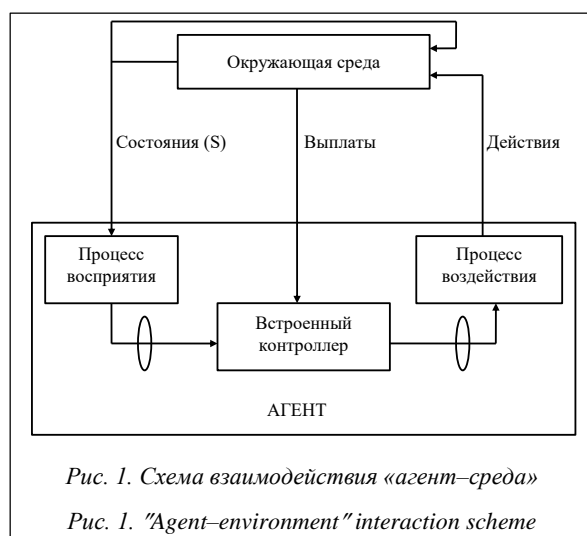


Рис. 1. Схема взаимодействия «агент–среда»

Рис. 1. "Agent-environment" interaction scheme

временном шаге t . Данный подход применим в прикладных задачах, когда завершающий шаг можно определить естественным образом исходя из природы решаемой задачи, то есть когда взаимодействие «агент–окружающая среда» можно разбить на последовательности, называемые *эпизодами*.

Основной проблемой RL-обучения является нахождение агентом компромисса между изучением и применением. Для получения большего вознаграждения агент должен предпочитать действия, ранее уже применявшиеся и показавшие свою эффективность с точки зрения получения поощрения. С другой стороны, чтобы обнаруживать такие действия, агенту необходимо пробовать выполнять новые действия. Таким образом, агент должен как применять уже известные действия, так и изучать новые для возможности иметь наилучший выбор в будущем. Важной характеристикой RL-обучения является получение отложенных вознаграждений, которые имеют место в сложных динамических системах. Это означает, что действие агента может повлиять не только на текущую награду, но и на все последующие.

В плане применения в ИС РВ TD-методы могут решать несколько задач: *задачу предсказания* значений некоторых переменных в течение нескольких временных шагов и *задачу управления*, основанную на RL-обучении агента тому, как влиять на окружающую среду. Таким образом, агент должен предсказывать последующие состояния окружающей среды и использовать эти значения для ее изменения с целью максимизации вознаграждений.

Для возможности обучения и адаптации к изменениям внешней среды агент должен обладать памятью для хранения предыстории. При этом возникает ряд проблем, связанных с объемом доступной агенту информации о прошлом, с запоминанием, хранением, использованием доступной информации и т.д. Для решения этих проблем агент может использовать скользящее окно для истории (наиболее простой метод) или строить зависящую от состояния прогнозную модель окружающей среды. Можно применить комбинацию этих подходов, когда агент анализирует чувствительную к предыстории политику непосредственно при обучении.

Несмотря на проблему поиска компромисса между изучением и применением, RL-обучение имеет ряд важных достоинств для применения в ИС РВ типа ИСППР РВ:

- использование простой обратной связи на основе скалярных платежей;
- поддержка режима оперативного реагирования, когда агенту необходимо быстро адаптироваться к изменениям внешней среды;
- интерактивность и возможность изменения (пополнения) анализируемых данных (предыстории);
- действенность в недетерминированных средах;

- эффективность в сочетании с темпоральными моделями для задач нахождения последовательных решений;
- открытость к модификации и сравнительная простота включения в интеллектуальные системы различного назначения (планирования, управления, обучения и т.д.).

Методы обучения с подкреплением на основе темпоральных различий

Рассмотрим RL-методы на основе темпоральных различий (TD-методы) в плане их применения в ИС РВ [1, 10]. TD-методы для решения задачи предсказания используют имеющийся опыт. При наличии некоторого опыта следования избранной стратегии TD-методы корректируют свои оценки, например, если имеет место посещение нетерминального состояния S_t в момент времени t , то корректируются оценки $V(S_t)$, основываясь на том, что случилось после этого посещения, то есть для корректировки оценки необходимо дождаться только следующего временного шага. Непосредственно в момент времени $t+1$ формируется целевое значение оценки и производится необходимая корректировка с учетом уже имеющегося вознаграждения r_{t+1} и оценки $V(S_{t+1})$.

Для наиболее простого TD-метода (метода TD(0)) справедливо $V_{S_t} \leftarrow V(S_t) + \alpha[r_{t+1} + \gamma V(S_{t+1}) - V(S_t)]$, где γ – ценность терминального состояния. При корректировке целью будет величина $r_{t+1} + \gamma V(S_{t+1})$. Так как в своих корректировках TD-метод частично основывается на существующих оценках, он является *самонастраивающимся*. Одним из преимуществ TD-методов является то, что они не требуют знания модели окружающей среды с ее вознаграждениями и вероятностным распределением последующих состояний.

Следует отметить, что при обучении на основе вероятностного метода Монте-Карло каждый раз необходимо ждать завершения эпизода, так как только тогда становится известной оценка (выгода), в то время как при использовании TD-методов необходимо дождаться лишь следующего временного шага. Данное преимущество TD-методов часто имеет решающее значение при использовании в ИС РВ, так как в некоторых ситуациях эпизоды могут быть настолько продолжительными, что задержки процесса обучения, связанные с необходимостью завершения эпизодов, будут слишком велики. Возможны также ситуации, когда возникают непрерывные задачи, а эпизоды как таковые отсутствуют.

TD-методы обучаются на основе каждого перехода вне зависимости от осуществляемых в дальнейшем действий и, соответственно, не чувствительны к ситуациям, когда необходимо игнорировать эпизоды или снижать значимость эпизодов, в которых предпринимаются экспериментальные

действия, что может сильно замедлить обучение. TD-методы в целом можно разделить на две основные категории – методы с интегрированной (on-policy) и методы с разделенной (off-policy) оценкой ценности стратегий. В методах с интегрированной оценкой стратегия, используемая для управления, аналогична оценочной стратегии, которая совершенствуется во время обучения. В методах с разделенной оценкой стратегия управления не имеет взаимосвязи с оценочной стратегией.

Метод SARSA – TD-метод с интегрированной оценкой ценности стратегий. Для данного метода необходимо оценить функцию $Q^{\pi}(s, a)$ для текущей стратегии π и для всех состояний s и действий a : $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]$, где α – постоянная длина шага; γ – ценность терминального состояния. Данная корректировка имеет место после каждого перехода из нетерминального состояния s_t . Если состояние s_{t+1} является терминальным (заключительным), то значение $Q(s_{t+1}, a_{t+1})$ полагается равным нулю. Это правило использует каждый элемент из пятерки $(s_t, a_t, r_{t+1}, s_{t+1}, a_{t+1})$, когда происходит переход от одной пары «состояние–действие» к другой. В методах с интегрированной оценкой ценности стратегий все время оценивается функция Q^{π} для стратегии поведения π и в то же время стратегия π делается более жадной по отношению к Q^{π} .

Свойство сходимости алгоритма SARSA непосредственно связано с зависимостью стратегии от функции Q . Например, можно использовать ϵ -жадную и ϵ -гибкую стратегии. Алгоритм сходится с вероятностью 1 к оптимальной стратегии и функции ценности действия при условии, что все пары «состояние–действие» посещались бесконечное число раз, и стратегия сходится в пределе к жадной стратегии (что может быть осуществлено, например, при помощи ϵ -жадных стратегий с $\epsilon = 1/t$).

Метод Q -обучения – метод с разделенной оценкой ценности стратегий, который находит оптимальные значения функции Q для выбора последующих действий и одновременно определяет оптимальную стратегию. Аналогично методу TD(0) в каждой итерации есть только знание о двух состояниях: s и одного из его предшествующих. Таким образом, значения функции Q позволяют получить некоторое представление о будущем качестве действий в предшествующих состояниях и сделать задачу выбора действия проще.

Для данного метода нужно оценить функцию ценности действия $Q^{\pi}(s, a)$ для текущей стратегии π и для всех состояний s и действий a , где эпизод состоит из последовательности перемежающихся состояний и пар «состояние–действие». Одношаговое Q -обучение характеризуется зависимостью $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)]$, где α – постоянная длина шага; γ – ценность терминального состояния. В этом случае искомая функция ценности действия Q непосред-

ственно аппроксимирует Q^* – оптимальную функцию ценности действия независимо от применяющейся стратегии. Стратегия определяет, какие пары «состояние–действие» посещаются и корректируются. В Q -обучении правило обновления всегда выполняется на основе жадной детерминированной стратегии, которая совершенствуется. И в то же время действия, выбранные для управления, основаны на другой стратегии (не зависящей от $Q(s, a)$). Например, для генерирования действий может быть использована стратегия с равномерным распределением по пространству действий.

Для обеспечения сходимости необходимо, чтобы все пары продолжали корректироваться. Это является минимальным требованием в том смысле, что каждый метод, гарантированно находящий оптимальную линию поведения, в общем случае должен удовлетворять данному условию. Установлено, что при таком условии и в случае стохастической аппроксимации для последовательности значений длины шага функция Q_t сходится к Q^* с вероятностью 1 [1].

Метод TD(λ) – метод, в котором временное различие имеет протяженность в n шагов. Вводится дополнительная переменная памяти, соответствующая каждому состоянию, – *след приемлемости*. Для состояния s в момент времени t след приемлемости обозначается $e^t(s)$. На каждом шаге следы приемлемости для всех состояний убывают с коэффициентом $\gamma\lambda$, а след приемлемости для посещаемого на данном шаге состояния увеличивается на 1: $e^t(s) = \gamma\lambda e_{t-1}(s)$, если $s \neq s_t$, $e^t(s) = \gamma\lambda e_{t-1}(s) + 1$, если $s = s_t$, где γ – ценность терминального состояния; λ – коэффициент затухания. Такие следы показывают степень приемлемости каждого состояния при происходящих изменениях в обучении, если возникает подкрепляющее событие. Таким образом, для метода TD(λ) имеем: $V_{s_t} \leftarrow V(s_t) + e(s_t)\alpha[r_{t+1} + \gamma V(s_{t+1}) - V(s_t)]$, $\forall s \in S : e(s) \neq 0$.

Используя следы приемлемости, все состояния должны быть обновлены на каждом шаге (выбор действия a в состоянии s_t и получение вознаграждения r в состоянии s_{t+1}). В то же время информация о текущих выплатах распространяется обратно к состояниям с более высокими значениями следов приемлемости. Можно показать, что при значении $\lambda = 0$ алгоритм становится аналогичным алгоритму TD(0), обновляя только состояния s_t на шаге $t+1$. При значении $\lambda = 1$ алгоритм поиска решения эквивалентен полному прогону метода, который имеет смысл только для эпизодических задач при оценке значений состояний после получения всех вознаграждений и подсчета полной выгоды.

Обучение с подкреплением в мультиагентной системе

Известно, что мультиагентная система – это группа автономных взаимодействующих между

собой субъектов (агентов), имеющих общую интерактивную среду и способных получать, хранить, обрабатывать и передавать информацию в интересах решения как собственных, так и корпоративных (общих для группы агентов) задач анализа и синтеза информации [9]. Структура мультиагентной системы для RL-обучения аналогична рассмотренной ранее схеме взаимодействия «агент–окружающая среда» (см. рис. 1) с отличием в том, что на окружающую среду оказывают влияние несколько агентов одновременно и, соответственно, действия каждого агента могут зависеть от действий остальных агентов системы.

К преимуществам мультиагентных систем в RL-обучении можно отнести следующие:

- возможность параллельных вычислений, так как используется распределенный характер взаимодействия агентов, в рамках ускорения работы системы;
- обмен опытом между агентами, средствами обучения и имитации, позволяющий помочь RL-агентам с похожими задачами обучаться быстрее и достичь более высокой производительности;
- отказоустойчивость – при выводе из строя одного или нескольких агентов система продолжает функционировать;
- масштабируемость – включение или исключение агента из системы не влияет на работу системы в целом.

Но при этом возникают определенные сложности:

- сложность задания цели обучения;
- нестационарность проблемы обучения, возникающая из-за того, что все агенты обучаются одновременно и каждый агент сталкивается с проблемой изменяющейся цели обучения, поэтому основная стратегия может меняться при изменении стратегий других агентов; таким образом, RL-агенту необходимо найти компромисс между использованием текущих знаний и исследованием среды для сбора информации и улучшения этих знаний;
- необходимость координации;
- экспоненциальный рост дискретного пространства состояний-действий, так как основной алгоритм Q -обучения оценивает значения всех возможных пар «состояние–действие», что ведет, соответственно, к экспоненциальному увеличению вычислительной сложности.

Реализация подсистемы прогнозирования для ИС РВ с включением методов обучения

На основе статистических и экспертных методов прогнозирования был предложен *комбинированный (интегрированный) метод прогнозирования* [10], который заключается в усреднении результатов, полученных на основе методов скользящей средней и Байеса [11] с учетом весовых коэффициентов. Затем полученный прогноз кор-

ректируется относительно значений ряда, полученного на основе метода экспоненциального сглаживания, а далее – с учетом экспертных методов ранжирования и непосредственной оценки. Вероятность каждого исхода, полученного статистическими методами, корректируется (увеличивается или уменьшается) в зависимости от значений экспертных оценок для указанных исходов.

Предложенная архитектура подсистемы прогнозирования (рис. 2) включает

- эмулятор, моделирующий состояние среды с использованием различных алгоритмов изменения параметров системы в оперативной БД;
- модуль прогнозирования на основе статистических методов (методов экстраполяции по скользящей средней, экспоненциального сглаживания и байесовского подхода) и прогнозирования на основе экспертных методов (методы ранжирования и непосредственной оценки);
- мультиагентный модуль RL-обучения, состоящий из группы независимых агентов, каждый из которых обучается на основе одного из разработанных TD-методов (TD(0), TD(λ), SARSA, Q-обучение), а также используемый для накопления знаний об окружающей среде и способный к адаптации, модификации и накоплению знаний;
- модуль принятия решений, предназначенный для анализа данных, поступающих от модулей прогнозирования, RL-обучения и принятия решений о последующих действиях, способе корректировки стратегий управления и т.д.

Выполнена программная реализация прототипа подсистемы прогнозирования с использованием статистического и экспертного модулей для решения задач экспертного диагностирования сложного технологического объекта – одной из подсистем АЭС (подсистема «1 контур» ВВЭР АЭС) с целью выполнения прогнозирования для оценки развития

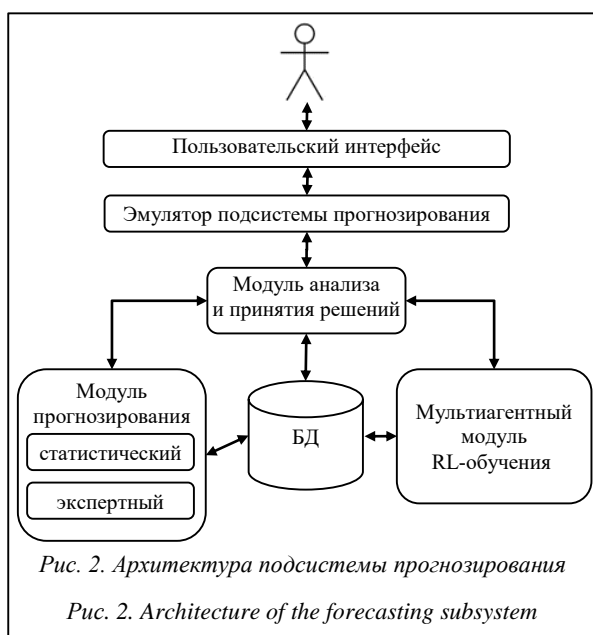


Рис. 2. Архитектура подсистемы прогнозирования

Рис. 2. Architecture of the forecasting subsystem

ситуации на объекте [10]. По результатам тестирования установлено, что необходимо привлечение дополнительных средств: методов RL-обучения на основе темпоральных различий, которые позволяют выявить имеющиеся закономерности посредством анализа предыстории процесса и таким образом уменьшить влияние случайных явлений.

Были разработаны и исследованы различные алгоритмы TD-методов (TD(0), TD(λ), SARSA, Q-обучение) с целью анализа возможности их применимости в интегрированной среде, а также для сравнения результатов прогнозирования при использовании различных методов и их комбинаций [10]. Задача исследования – проектирование мультиагентной системы RL-обучения и ее интеграция в подсистему прогнозирования, а также нахождение наиболее предпочтительных для включения в состав ИС РВ типа ИСППР РВ методов RL-обучения и прогнозирования и оценка эффективности функционирования мультиагентных систем в рамках ИСППР РВ.

Заключение

В работе были проанализированы различные методы RL-обучения и реализованы соответствующие алгоритмы в плане их последующей интеграции в блок прогнозирования для ИС РВ типа ИСППР РВ. Особое внимание уделено методам на основе темпоральных различий (TD-методам). Предложен комбинированный метод прогнозирования, основанный на статистических и экспертных методах прогнозирования, и реализованы алгоритмы для комбинированного метода. Предложена архитектура подсистемы прогнозирования, включающая модуль прогнозирования, мультиагентный модуль RL-обучения и модуль анализа и принятия решений.

В настоящее время разрабатывается мультиагентный модуль RL-обучения для его включения в интегрированную среду, ориентированную на использование в ИСППР РВ семиотического типа, с целью расширения области применения, повышения производительности и эффективности функционирования современных ИСППР РВ.

Работа выполнена при финансовой поддержке РФФИ (проекты №№ 17-07-00553, 16-51-00058) и проекта по государственному заданию № 2.737.2014/К.

Литература

1. Саттон Р.С., Барто Э.Г. Обучение с подкреплением; [пер. с англ.]. М.: БИНОМ. Лаборатория знаний, 2011. 400 с.
2. Вагин В.Н., Еремеев А.П. Некоторые базовые принципы построения интеллектуальных систем поддержки принятия решений реального времени // Изв. РАН. Теория и система управления. 2001. № 6. С. 114–123.
3. Shani G., Brafman R.I., Shimony S.E. Model-based online learning of POMDPs. Proc. Europ. Conf. on Machine Learning, 2005, pp. 353–364.
4. Рыбина Г.В., Паронджанов С.С. Технология построения динамических интеллектуальных систем: учеб. пособие. М.: Изд-во НИЯУ МИФИ, 2011. 240 с.

5. Осипов Г.С. Методы искусственного интеллекта. М.: Физматлит, 2011. 296 с.
6. Busoniu L., Babuska R., and De Schutter B. Multi-agent reinforcement learning: An overview. Chapter 7 in *Innovations in Multi-Agent Systems and Applications-1* (D. Srinivasan and L.C. Jain, eds.), vol. 310 of *Studies in Computational Intelligence*, Berlin, Germany: Springer, 2010, pp. 183–221.
7. Doshi-Velez F., Pineau J., Roy N. Reinforcement learning with limited reinforcement: Using Bayes risk for active learning in POMDPs, *Artificial Intelligence*, 2012, no. 187, pp. 115–132.
8. Еремеев А.П., Подогов И.Ю. Обобщенный метод иерархического подкрепленного обучения для интеллектуальных систем поддержки принятия решений // Программные про-

дукты и системы. 2008. № 2. С. 35–39.

9. Sort J., Singh S., Lewis R.L. Variance-based rewards for approximate Bayesian reinforcement learning. *Proc. Conf. Uncertainty in Artificial Intelligence*, 2010, pp. 564–571.
10. Еремеев А.П., Кожухов А.А. Разработка интегрированной среды на основе методов прогнозирования и обучения с подкреплением для интеллектуальных систем реального времени // IS&IT'16: тр. Конгресса по интелект. сист. и информ. технологиям. Науч. изд. в 3 т. Таганрог: Изд-во ЮФУ, 2016. Т. 1. С. 140–149.
11. Ross S., Chaib-draa B., Pineau J. Bayes-adaptive POMDPs. *Proc. Conf. Advances in Neural Information Processing Systems* 20, 2007, pp. 1225–1232.

Software & Systems

DOI: 10.15827/0236-235X.117.028-033

Received 17.11.16

2017, vol. 30, no. 1, pp. 28–33

IMPLEMENTATION OF REINFORCEMENT LEARNING METHODS BASED ON TEMPORAL DIFFERENCES AND A MULTI-AGENT APPROACH FOR REAL-TIME INTELLIGENT SYSTEMS

A.P. Eremeev¹, Dr.Sc. (Engineering), Professor, eremeev@appmat.ru

A.A. Kozhukhov¹, Postgraduate Student, saanchezz@yandex.ru

¹ National Research University "MPEI", Krasnokazarmennaya St. 14, Moscow, 111250, Russian Federation

Abstract. The paper describes implementation of reinforcement learning methods based on time (temporal) differences and a multi-agent technology. The authors examine the possibilities of combining learning methods with statistical and expert methods of forecasting for further integration into an instrumental software environment to use in modern and advanced real-time intelligent systems (RT IS), a type of real-time intelligent decision support systems (RT IDSS).

There is an analysis of reinforcement learning (RL-learning) methods in terms of using them in RT IS, main components, benefits and tasks. The paper focuses on the methods of RL-learning based on time (temporal) differences (TD-methods) and presents the developed corresponding algorithms. The authors consider the possibility of including RL-learning methods into a multi-agent environment and combining them with statistical and expert forecasting methods in terms of integration into the environment, which was developed for RT IDSS for complex technical object control and diagnosis.

The paper proposes the architecture of the forecasting subsystem prototype consisting of an emulator, which simulates the state of environment, forecasting module, analysis and decision-making module and a multi-agent RL-learning module. There is software implementation of the forecasting subsystem prototype using a multi-agent approach in order to solve the problem of the complex technological object expert diagnosis.

According to the results of testing and validation of the developed system, the paper considers the conclusions about the efficiency and expediency of including into the RT IDSS.

Keywords: artificial intelligence, intelligent system, real time, reinforcement learning, forecasting, decision support, program tools.

Acknowledgements. The work has been financially supported by RFBR (projects no. 17-07-00553, 16-51-00058) and a project by the state order no. 2.737.2014/K.

References

1. Sutton R.S., Barto A.G. *Reinforcement Learning*. London, MIT Press, 2012, 320 p. (Russ. ed.: Moscow, BINOM Publ., 2011, 400 p.).
2. Vagin V.N., Eremeev A.P. Some basic construction principles of real-time intelligent decision support systems. *Izv. RAN. Teoriya i sistemy upravleniya* [Journal of Computer and Systems Sciences International]. 2001, no. 6, pp. 114–123 (in Russ.).
3. Shani G., Brafman R.I., Shimony S.E. Model-based online learning of POMDPs. *Proc. European Conf. on Machine Learning*. 2005, pp. 353–364.
4. Rybina G.V., Parondzhanov S.S. *Tekhnologia postroeniya dinamicheskikh intellektualnykh sistem* [The technology of building dynamic intelligent systems]. Moscow, MEPHI Publ., 2011, 240 p.
5. Osipov G.S. *Metody iskusstvennogo intellekta* [Methods of artificial intelligence]. Moscow, Fizmatlit Publ., 2011, 296 p.
6. Busoniu L., Babuska R., De Schutter B. *Multi-agent reinforcement learning*. Berlin, Germany, Springer Publ., 2010, ch. 7., vol. 310, pp. 183–221.
7. Doshi-Velez F., Pineau J., Roy N. Reinforcement learning with limited reinforcement: Using Bayes risk for active learning in POMDPs. *Artificial Intelligence*. 2012, vol. 187–188, pp. 115–132.
8. Eremeev A.P., Podogov I.U. Generalized method of hierarchical reinforcement learning for intelligent decision support systems. *Programmnye produkty i sistemy* [Software and Systems]. 2008, no. 2, pp. 35–39 (in Russ.).
9. Sort J., Singh S., Lewis R.L. Variance-based rewards for approximate Bayesian reinforcement learning, *Proc. Uncertainty in Artificial Intelligence*. 2010, pp. 564–571.
10. Eremeev A.P., Kozhukhov A.A. Analysis and development of reinforcement learning methods based on temporal differences for real time intelligent systems. *Proc. 15th National Conf. on Artificial Intelligence with International Participation KII-2016*. Vol. 1, Smolensk, Universum Publ., 2016, pp. 323–330 (in Russ.).
11. Ross S., Chaib-draa B., Pineau J. Bayes-adaptive POMDPs. *Advances in Neural Information Processing Systems*. 2008, vol. 20, pp. 1225–1232.

УДК 51-74

DOI: 10.15827/0236-235X.117.034-039

Дата подачи статьи: 19.08.16

2017. Т. 30. № 1. С. 34–39

МОДЕЛИРОВАНИЕ ВОЗДЕЙСТВИЯ АТАКИ BLACK HOLE НА БЕСПРОВОДНЫЕ СЕТИ

В.В. Шахов, к.ф.-м.н., доцент, старший научный сотрудник, shakhov@rav.sccc.ru;

А.Н. Юргенсон, к.ф.-м.н., научный сотрудник, nastya@rav.sccc.ru;

О.Д. Соколова, к.т.н., старший научный сотрудник, olga@rav.sccc.ru

*(Институт вычислительной математики и математической геофизики СО РАН,
просп. Академика Лаврентьева, 6, г. Новосибирск, 630090, Россия)*

Технологии, основанные на беспроводных сенсорных сетях, проникают в самые важные сферы жизнедеятельности общества. Многие решения в области архитектуры Интернета вещей опираются на результаты исследований беспроводных сенсорных сетей, в частности, это касается предложений, разработанных в рамках ряда проектов Седьмой рамочной программы Европейского союза по развитию научных исследований и технологий. Следовательно, особое внимание необходимо уделять обеспечению безопасности таких сетей.

В статье обсуждаются проблемы функционирования сетей в условиях несанкционированных вторжений. Обеспечить абсолютную защиту, полностью нивелировать последствия вторжений возможно далеко не во всех случаях. Однако эффективный выбор механизмов защиты позволит существенно снизить ущерб. Для этого необходимо разрабатывать и исследовать адекватные математические модели.

Авторы рассматривают моделирование атаки Black Hole на узлы беспроводных сенсорных сетей и исследуют оценку нанесенного ущерба. Эта атака является одним из наиболее опасных разрушающих информационных воздействий, в результате ее может теряться более 90 % информации, передаваемой в сток. В качестве модели беспроводной сети используются графы единичных кругов (UDG-графы), которые наиболее адекватно описывают связи в этих сетях, где передача информации между узлами возможна, если они находятся в пределах взаимной достижимости радиосигнала.

Для моделирования передачи данных по выбранному алгоритму маршрутизации в графе строится остовное дерево. Авторами получены формулы для вычисления аналитических оценок для некоторых случаев вида остовного дерева. Чтобы оценить уязвимость дерева передачи данных к атакам, использовалась величина «нормированное число вершин, от которых потеряна информация» – среднее число вершин, от которых потеряна информация, деленное на общее число вершин в дереве.

Полученные аналитические результаты согласуются с результатами имитационного моделирования. Предложен метод противодействия атакам типа Black Hole, оценена его эффективность.

Ключевые слова: беспроводные сенсорные сети, безопасность, атака Black Hole.

Новейшие достижения в области сетевых технологий, физики полупроводников и материаловедения позволили приступить к повсеместной разработке и внедрению *беспроводных сенсорных сетей* (БСС). Первоначально научно-исследовательские и конструкторские работы по данной теме проводились при поддержке Агентства по перспективным научно-исследовательским разработкам при министерстве обороны США (Defense Advanced Research Projects Agency, DARPA). Однако в настоящее время технологии БСС находят применение в самых разных сферах человеческой жизнедеятельности: изучение биологии диких животных и птиц, обнаружение лесных пожаров и наводнений, мониторинг загрязнения воздуха, системы «умный дом», системы предупреждения техногенных аварий, контроль состояния пожилых людей и пациентов госпиталя, новейшие методы медицинской диагностики, отслеживание транспортных потоков, промышленные робототехнические системы, современные сельскохозяйственные технологии и т.д. Интенсивные исследования по проблематике БСС проводятся не только ведущими мировыми научными центрами, но и коммерческими компаниями, такими как IBM, Intel, Samsung, Cisco Systems, Google и др. Для координации работ в об-

ласти БСС создан альянс ZigBee, куда вошли крупнейшие разработчики аппаратных и программных средств. Результатом усилий альянса стала спецификация протоколов сетевого и прикладного уровня, разработанная на основе стандарта IEEE 802.15.4, описывающего физический уровень и нижний канальный подуровень (управление доступом к среде) для низкоскоростных беспроводных персональных сетей [1]. Недавно альянс анонсировал единый стандарт ZigBee 3.0, объединяющий лидирующие на рынке беспроводные стандарты, позиционируя его как удобное средство для разработчиков продуктов и услуг, относящихся к Интернету вещей (Internet of Things, IoT).

Следует отметить, что развитию Интернета вещей уделяется особое внимание во многих странах. Данный рынок является очень перспективным: по оценкам специалистов Cisco Systems, число устройств, подключенных к Интернету, уже составляет десятки миллиардов. Международные консалтинговые компании, специализирующиеся на рекомендациях по стратегическому управлению, прогнозируют в ближайшие несколько лет крупный экономический эффект от развития Интернета вещей. Фондом развития интернет-инициатив (Российским фондом венчурных инвестиций)

объявлено о создании консорциума для формирования российского пакета технологий Интернета вещей в партнерстве с инвестиционным холдингом GS Group и операторами сотовой связи. Многие решения в области архитектуры IoT опираются на результаты исследований БСС [2], в частности, предложения, разработанные в рамках проектов Седьмой рамочной программы Европейского союза по развитию научных исследований и технологий [3].

Таким образом, технологии, основанные на БСС, проникают в самые важные сферы жизнедеятельности общества, следовательно, особое внимание необходимо уделять вопросам безопасности указанных сетей. Из-за ограничений, налагаемых требованиями рынка на компоненты БСС, и особенностей функционирования БСС указанные сети легко подвергаются атакам. Обеспечить абсолютную защиту БСС, полностью нивелировать последствия несанкционированных вторжений возможно далеко не во всех случаях. Однако эффективный выбор механизмов защиты позволит существенно снизить ущерб. Для этого необходимо разрабатывать и анализировать соответствующие математические модели.

В данной статье дано общее описание БСС, рассмотрены причины их уязвимости и некоторые атаки. Рассмотрены вопросы моделирования БСС. Предлагаются подходы к моделированию атаки Black Hole, к оценке нанесенного сети ущерба, рассматривается способ противодействия данной атаке. Анализ эффективности предлагаемого механизма противодействия атаке и выводы завершают статью.

Беспроводные сенсорные сети

БСС образованы большим количеством сетевых узлов [4], называемых мотами, – миниатюрных автономных устройств, способных собирать информацию с территории в определенном радиусе действия и передавать ее другим устройствам. Каждое такое устройство содержит модуль сбора данных (температуры, давления, освещенности и т.д.) и автономный источник питания. Также каждый мот оснащен радиотрансивером или другим устройством беспроводной связи, то есть данные передаются в сети по радиоканалу. Для аккумуляции всей собираемой информации сеть содержит мощный узел (сток, базовая станция), подключенный к стационарному источнику питания. Данные собираются в этот сток по определенному алгоритму маршрутизации. Объединенные в беспроводную сеть, все узлы образуют распределенную самоорганизующуюся систему сбора и передачи информации. Преимущества систем на основе сенсорных сетей – возможность развертывания в труднодоступных местах, беспроводная связь, самоорганизация (возможность перераспределения маршрутов в случае выхода из строя некоторых узлов).

Несмотря на очевидные преимущества систем с беспроводной связью, они отличаются и большей по сравнению с проводными сетями уязвимостью. Для обеспечения отказоустойчивости БСС необходимо решить ряд проблем, возникающих вследствие обмена информацией в открытой распределенной самоорганизующейся системе, топология которой может изменяться во времени [5].

Основные причины уязвимости БСС: доступность среды передачи, незащищенность узлов, относительная невозможность анализа всего трафика на предмет обнаружения аномалий, невозможность использования криптографии и сложного математического аппарата из-за ограниченности ресурсов. Именно ограниченность ресурсов сенсора позволяет легко вывести его из строя или использовать по усмотрению злоумышленника. Отказы узлов могут возникать в случае как несанкционированных вторжений в сеть, так и сбоя легальных протоколов. Выход из строя всего одного сенсора может привести к тому, что теряются потоки данных от множества других сенсоров, использующих атакующий узел в качестве промежуточного на пути к стоку. Следовательно, ущерб в этом случае будет нанесен значительному сегменту сети.

Для организации разрушающего воздействия используются радиопомехи, вредоносные программы с целью перехвата информации, перевод узла в спящий режим [6] и др. Например, целью атаки Node replication (клонирование узла) является фальсификация данных, передаваемых в сток. Атака Jamming (создание помех) оказывает воздействие на каналы и затрудняет передачу информации. Атака Black Hole (черная дыра) использует уязвимость протоколов маршрутизации БСС: атакующий узел посылает соседним узлам информацию о том, что он находится близко к стоку, вследствие чего маршрутизация меняется, потоки данных проходят через этот узел, далее информация блокируется. Если сеть обладает способностью к самовосстановлению, то есть протоколы передачи данных позволяют обнаруживать неисправные узлы и исключать их из маршрутов, эффект от указанных разрушающих воздействий не будет продолжительным. Однако существуют и более успешные способы организации атак [7, 8].

Для противодействия атакам необходимо повышать надежность используемых протоколов маршрутизации, обеспечивать мобильность стоков (возможность замены стока) или принимать другие меры. Отсюда вытекает необходимость умения моделировать работу сети в различных режимах, особенно под влиянием воздействий, чтобы определить оптимальный уровень защиты.

Моделирование сенсорных сетей

БСС удобно моделировать графом, в котором вершины распределены случайным образом на об-

ласти с евклидовой метрикой. Так как сигнал от каждого узла распространяется во все стороны, место, где сигнал может быть получен другим узлом, моделируется кругом. Один узел может передавать информацию другому, если они находятся в пределах взаимной достижимости сигнала. Следовательно, две вершины графа соединяются ребром, если одна вершина находится в круге, образованном другой вершиной. Если все узлы имеют передатчики одинаковой мощности, круги имеют один и тот же радиус. Это означает, что в моделируемом графе ребро между двумя вершинами существует, если расстояние между ними в евклидовой метрике меньше заданного числа либо равно ему. В этом случае в качестве модели удобно использовать класс графов, которые называются Unit Disk Graphs (UDG-графы).

Определение [9]. Граф $G=(V, E)$ называется UDG-графом (unit disk graph, граф единичных кругов), если ребро $e=(u, v)$ между вершинами $u, v \in V$ существует только в том случае, когда в евклидовой метрике расстояние между u и v меньше либо равно 1.

Рассмотрим граф единичных кругов $G=(V, E)$, $|V|=n$, в котором всем ребрам $e \in E$ приписаны некоторые веса $f(e)$, зависящие от их длины. Например, потребление энергии для передачи данных от одного узла к другому пропорционально квадрату расстояния между ними. Надежность соединения, а значит, и количество повторных передач также зависят от расстояния [10]. В графе G выделяем одну вершину s – сток, то есть узел, в котором собирается вся информация, передаваемая вершинами сети. Остальные вершины могут принимать и передавать информацию для отправки ее в сток. Существует множество алгоритмов передачи данных в сенсорных сетях, которые оптимизируют различные показатели, например, количество потребляемой энергии, скорость передачи данных и др. На основе таких алгоритмов разрабатываются протоколы маршрутизации, по которым осуществляются сбор информации со всех узлов сети и передача ее в сток. Для построения маршрутов от каждой вершины к стоку на графе строится остовное дерево T (то есть дерево, содержащее все вершины графа) с направленными ребрами. Алгоритм построения такого дерева зависит от выбранного алгоритма маршрутизации [11, 12]. По ребрам дерева T происходит передача данных от каждой вершины v_i в сток s .

Количество направленных дуг, входящих в вершину v_i , будем обозначать $d^+(v_i)$. Степенью вершины $d(v_i)$ будем называть число входящих и исходящих из нее дуг.

Атака Black Hole

Одним из наиболее опасных разрушающих воздействий в БСС является атака Black Hole. В ре-

зультате действия атак этого типа может теряться более 90 % информации, передаваемой в сток [13]. Атаку можно организовать двумя способами. Один способ – размещение злоумышленником в области действия сети нового узла, с помощью которого в дальнейшем организуется атака. Воздействия такого рода относительно легко обнаруживаются и локализуются стандартными механизмами БСС. Более опасным является другой способ, когда осуществляется взлом одного из легальных узлов, уже участвующих в информационном обмене.

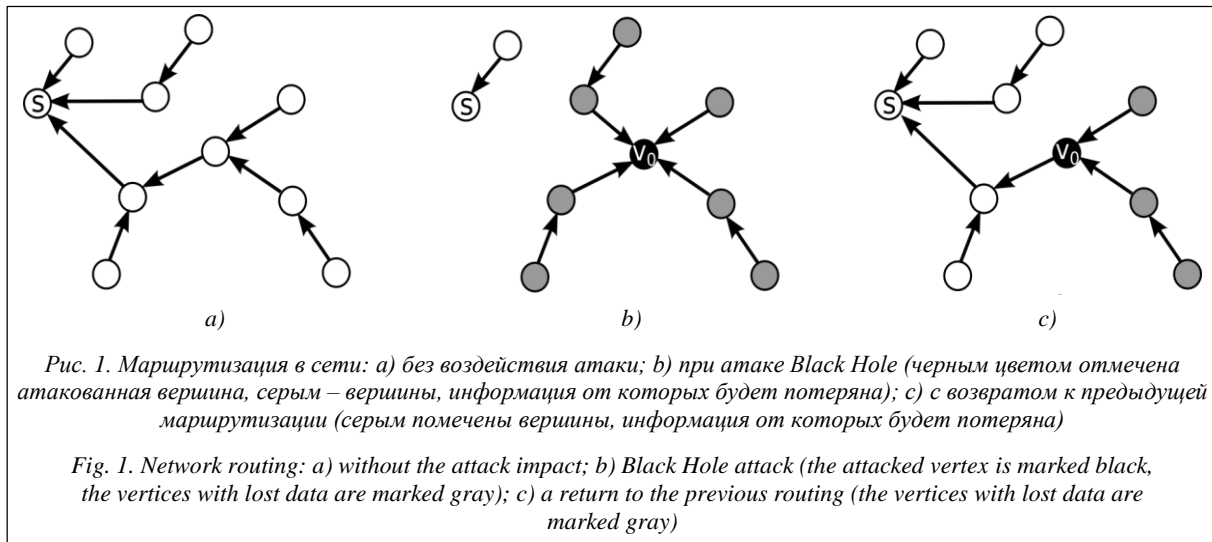
Контролируемый злоумышленником узел удаляет все пакеты, переданные в него другими узлами для транзитной передачи. Кроме того, взломанный узел v_0 может распространять по сети информацию, что он является ближайшим узлом к стоку s , вследствие чего самоорганизующаяся сеть, каковой является БСС, меняет маршрутизацию, и остальные узлы, находящиеся ближе к v_0 , чем к s , передают в v_0 свои пакеты для дальнейшей передачи в s .

В основе предлагаемого авторами метода защиты от атаки Black Hole лежит следующая идея. Так как в результате атаки собранная в узле v_0 информация блокируется, интенсивность транзитного трафика и нагрузки на сток s снижается. Основываясь на наблюдениях показателей трафика и используя методы обнаружения разладки случайных процессов, можно обнаружить несанкционированное вторжение. Как только возникает подозрение на наличие атаки, всем узлам отправляется команда вернуться к прежнему выбору транзитных узлов на пути к стоку. Данный сигнал может, например, передаваться мощным передатчиком, интегрированным со стоком, сразу для всех узлов сети. В случае использования злоумышленником нового узла эффект атаки полностью нивелируется. Если же для атаки использовался узел, участвовавший ранее в маршрутизации, восстановление маршрутов позволяет снизить потери, так как пропадает только информация, передаваемая транзитом через v_0 (ситуация показана на рисунке 1).

Оценим эффективность данного способа защиты. В остовном дереве T , построенном в графе G , множество узлов, информация от которых потеряна, образует подмножество $V' \subseteq V$. Таким образом, для оценки устойчивости дерева T к атаке удобно взять в качестве параметра количество таких узлов $n' = |V'|$ или нормированное количество: n' , деленное на общее число вершин в графе $n = |V|$.

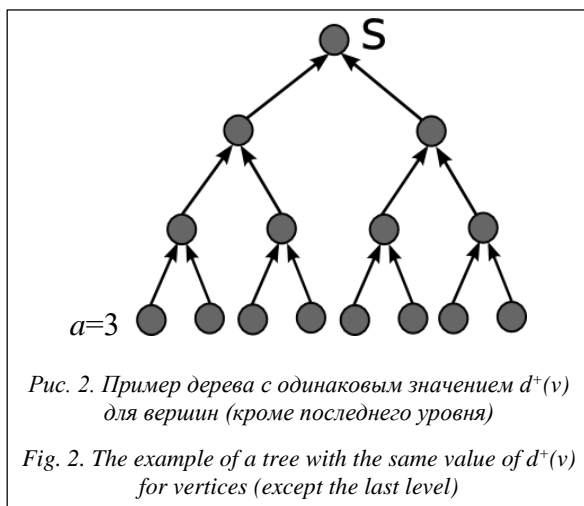
Для некоторых простых случаев вида остовного дерева T можно вычислить аналитические оценки среднего числа вершин, от которых теряется информация. Например, простым случаем можно считать дерево, у которого каждая вершина имеет одинаковое число потомков.

Рассмотрим дерево T , имеющее n вершин, для каждой вершины v значение $d^+(v) = k$, кроме вер-



шин последнего уровня. Считаем, что каждая ветвь дерева имеет одинаковое число уровней.

Пронумеруем уровни в дереве: сток считаем нулевым уровнем; вершины, передающие информацию в сток, – первый уровень и т.д., последний уровень имеет номер a (рис. 2). Тогда величины a и k связаны следующим соотношением: $k^{a+1} = n(k-1)+1$.



Считаем, что каждая вершина в дереве с равной вероятностью может подвергнуться атаке. Отсюда следует, что математическое ожидание числа вершин n' , от которых теряется информация под воздействием атаки Black Hole, равно:

$$E(n') = \frac{1}{n} \left(n + \sum_{j=1}^a k^j \sum_{i=1}^{a-j} k^i \right) = a + 1 - \frac{1 - \frac{a+1}{k}}{k-1}. \quad (1)$$

То есть чем больше число k , тем меньше $E(n')$.

В общем случае, когда нет условия о равенстве величины $d^+(v)$ для всех вершин и отсутствует ограничение на количество уровней, математическое ожидание числа вершин n' , от которых теряется информация под воздействием атаки Black Hole, будет следующим:

$$E(n') = \frac{1}{n} (1 + 2 |V_1| + \dots + (a+1) |V_a|), \quad (2)$$

где $V_j \subseteq V$ – вершины, принадлежащие уровню j .

Из формулы (2) следует, что, чем меньше число транзитов для передачи данных от вершины к стоку (число хопов), тем меньше $E(n')$.

Величину $E(n')$ можно интерпретировать как средний номер уровня для вершин (то есть сумма номеров уровней всех вершин, деленная на количество вершин). Таким образом получаем приближенную оценку количества атакованных узлов.

Выводы

Исследование воздействия атаки Black Hole в общем случае проведено с помощью имитационного моделирования. Для генерации случайных UDG-графов, отображающих топологию БСС, использовался подход, описанный в работе [14]. В качестве алгоритма маршрутизации выбран Minimum Energy Route [15]. Для оценки уязвимости дерева передачи данных к атакам использовалась величина «нормированное число вершин, от которых потеряна информация» – среднее число вершин, от которых потеряна информация, деленное на общее число вершин в дереве. Так, для БСС, содержащей 500 узлов, количество атакованных узлов не превышает 10 %, а для БСС, содержащей 2 000 узлов, из строя будет выведено примерно 5 % узлов. Полученные в результате имитационного моделирования оценки хорошо согласуются с формулами, выведенными в предыдущем разделе.

Заметим, что ущерб от разрушающего воздействия атаки Black Hole на узлы беспроводной сенсорной сети существенно зависит от того, какой алгоритм маршрутизации применяется в сети для сбора данных. За счет выбора надлежащего алгоритма можно существенно повысить надежность и живучесть БСС. Данное направление является темой будущих исследований авторов.

Работа выполнена при поддержке РФФИ, грант № 14-07-00769 а.

Литература

1. Oliveira T., Godoy E. ZigBee wireless dynamic sensor networks: feasibility analysis and implementation guide. *IEEE Sensors Jour.*, 2016, vol. 16, iss. 11, pp. 4614–4621.
2. McEwen A., Cassimally H. Designing the Internet of Things. John Wiley & Sons Ltd, UK, 2014, 336 p.
3. Uckelmann D., Harrison M., Michahelles F. Architecting the Internet of Things. Springer-Verlag Berlin Heidelberg, 2011, 356 p.
4. Akyildiz I.F., Su W., Sankarasubramaniam Y., Cayirci E. Wireless sensor networks: a survey. *Comp. Networks*, 15 March 2002, vol. 38, iss. 4, pp. 393–422.
5. Иващенко А.В., Минаев А.А., Сподобаев М.Ю. Шаблон агента-медиатора для программного обеспечения сенсорных сетей // Программные продукты и системы. 2015. № 3. С. 166–170.
6. Shakhov V.V., Popkov V.K. Performance analysis of sleeping attacks in wireless sensor networks. *Proc. IEEE Region 8th Intern. Conf. on Computational Technologies in Electrical and Electronics Engineering (SIBIRCON)*, 2008, pp. 418–420.
7. Cayirci E., Rong C. Security in Wireless Ad Hoc and Sensor Networks. John Wiley & Sons, 2009, 280 p.
8. Praveen K.S., Gururaj H.L., Ramesh B. Comparative analysis of Black Hole attack in ad hoc network using AODV and OLSR protocols. *Procedia Comp. Sc.*, 2016, vol. 85, pp. 325–330.
9. Clark A., Colbourn C., Johnson D. Unit disk graphs. *Discrete Mathematics*, 1990, vol. 86, pp. 165–177.
10. Shakhov V. Experiment design for parameter estimation in sensing models. *Springer LNCS*, 2013, vol. 8072, pp. 151–158.
11. Singh S.K., Singh M.P., Singh D.K. Routing protocols in wireless sensor networks – a survey. *IJCSSES*, 2010, vol. 1, no. 2, pp. 63–83.
12. Safonov A., Lyakhov A., Urgenson A., Sokolova O. Wireless groupcast routing with palette of transmission methods. *Multiple Access Communications*, 2012, pp. 97–108.
13. Dokurer S., Erten Y., Acar C. Performance analysis of ad-hoc networks under black hole attacks. *Proc. of IEEE Int. Conf. Southeast*, March, 2007, pp. 148–153.
14. Shakhov V.V., Sokolova O., Yurgenson N. A fast method for network topology generating. *Lecture notes in comp. sc.*, Springer, 2014, vol. 8715, pp. 96–101.
15. Yang L., Yang H.C. and Wu K. Minimum-energy route configuration for wireless ad hoc networks. *2006 IEEE Intern. Performance Comp. and Communications Conf.*, Phoenix, AZ, 2006, pp. 6–14.

Software & Systems

DOI: 10.15827/0236-235X.117.034-039

Received 19.08.16

2017, vol. 30, no. 1, pp. 34–39

MODELLING AND SIMULATION OF BLACK HOLE ATTACK ON WIRELESS NETWORKS

V.V. Shakhov¹, Ph.D. (Physics and Mathematics), Associate Professor, Senior Researcher, shakhov@rav.sgcc.ru

A.N. Yurgenson¹, Ph.D. (Physics and Mathematics), Research Associate, nastya@rav.sgcc.ru

O.D. Sokolova¹, Ph.D. (Engineering), Senior Researcher, olga@rav.sgcc.ru

¹ Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Academician Lavrentev Ave. 6, Novosibirsk, 630090, Russian Federation

Abstract. The technologies based on wireless sensor networks can be used in a wide range of vital applications. There are several implementations of the Internet of Things architecture based on wireless sensor networks. For example, the core objective in the projects of the 7th Framework Programme funded by the European Union was to provide the technical foundation for WSN technology in IoT products and services. As wireless sensor networks based applications are deployed, security becomes an essential requirement.

In this paper the authors discuss the state-of-arts for security issues in WSN. It is impossible in all cases to provide absolute protection and eliminate the consequences of intrusion. However, the effective range of protection mechanisms will significantly reduce the damage. To achieve this it is necessary to develop and explore appropriate mathematical models.

The paper focuses on the attack named Black Hole. This attack has one of the most dangerous destructive information impacts. As a result, more than 90 % of the information transmitted to the sink may be lost. The direct transmission of information between the nodes in WSN is possible if they are within each other's radio reachability. Therefore, the unit disk graphs (UDG-graphs) might be used as a wireless network model. Communication in these networks are described by UDG-models the most appropriate. To simulate data transmission by the routing algorithm in the graph, a spanning tree is constructed. The authors have obtained the formula for calculating analytical estimates for some cases of a spanning tree structure. To assess the vulnerability of this tree to attacks the authors used the value “normalized number of vertices with lost information”. It shows the average number of vertices which lose information, divided by the total number of nodes on the tree. The analytical results are consistent with simulation results. The paper offers a counteracting method against Black Hole and provides the corresponding performance analysis as well.

Keywords: wireless sensor networks, security, Black Hole attack.

Acknowledgements. The research has been financially supported by RFBR, grant no. 14-07-00769 а.

References

1. Oliveira T., Godoy E. ZigBee Wireless Dynamic Sensor Networks: Feasibility Analysis and Implementation Guide. *IEEE Sensors Jour.* 2016, vol. 16, iss. 11, pp. 4614–4621.

2. McEwen A., Cassimally H. *Designing the Internet of Things*. John Wiley & Sons Publ., UK, 2014.
3. Uckelmann D., Harrison M., Michahelles F. *Architecting the Internet of Things*. Springer-Verlag Berlin Heidelberg Publ., 2011.
4. Akyildiz I.F., Su W., Sankarasubramaniam Y., Cayirci E. Wireless sensor networks: a survey. *Computer Networks*. 2002, vol. 38, iss. 4, pp. 393–422.
5. Ivaschenko A.V., Minaev A.A., Spodobaev M.Yu. A mediator pattern for sensor networks software. *Programmnye produkty i sistemy* [Software & Systems]. 2015, no. 3, pp. 166–170 (in Russ.).
6. Shakhov V.V., Popkov V.K. Performance analysis of sleeping attacks in wireless sensor networks. *Proc. IEEE Region 8 Int. Conf. on Computational Technologies in Electrical and Electronics Engineering (SIBIRCON)*. 2008, pp. 418–420.
7. Cayirci E., Rong C. *Security in Wireless Ad Hoc and Sensor Networks*. John Wiley & Sons Publ., 2009.
8. Praveen K.S., Gururaj H.L., Ramesh B. Comparative Analysis of Black Hole Attack in Ad Hoc Network Using AODV and OLSR Protocols. *Procedia Computer Science*. 2016, vol. 85, pp. 325–330.
9. Clark A., Colbourn C., Johnson D. Unit disk graphs. *Discrete Mathematics*. 1990, vol. 86, pp. 165–177.
10. Shakhov V. Experiment Design for Parameter Estimation in Sensing Models. *Springer LNCS*. 2013, vol. 8072, pp. 151–158.
11. Singh Sh.K., Singh M.P., Singh D.K. Routing Protocols in Wireless Sensor Networks – A Survey. *Int. Jour. of Computer Science & Engineering Survey (IJCSES)*. 2010, vol. 1, no. 2, pp. 63–83.
12. Safonov A., Lyakhov A., Yurgenson A., Sokolova O. Wireless groupcast routing with palette of transmission methods. *Multiple Access Communications*. 2012, pp. 97–108.
13. Dokurer S., Erten Y., Acar C. Performance analysis of ad-hoc networks under black hole attacks. *Proc. of IEEE Int. Conf. SoutheastCon 2007*. 2007, pp. 148–153.
14. Shakhov V.V., Sokolova O., Yurgenson N. A Fast Method for Network Topology Generating. *Lecture Notes in Computer Science*. Springer Publ., 2014, vol. 8715, pp. 96–101.
15. Yang L., Yang H.C., Wu K. Minimum-energy route configuration for wireless ad hoc networks. *2006 IEEE Int. Performance Computing and Communications Conf.* Phoenix, AZ, 2006, pp. 6–14.

Примеры библиографического описания статьи

1. Шахов В.В., Юргенсон А.Н., Соколова О.Д. Моделирование воздействия атаки Black Hole на беспроводные сети // Программные продукты и системы. 2017. Т. 30. № 1. С. 34–39; DOI: 10.15827/0236-235X.117.034-039.
2. Shakhov V.V., Yurgenson A.N., Sokolova O.D. Modelling and simulation of Black Hole attack on wireless networks. *Programmnye produkty i sistemy* [Software & Systems]. 2017, vol. 30, no. 1, pp. 34–39 (in Russ.); DOI: 10.15827/0236-235X.117.034-039.

УДК 519.816
DOI: 10.15827/0236-235X.117.040-044

Дата подачи статьи: 25.08.16
2017. Т. 30. № 1. С. 40–44

ПРОГНОЗИРОВАНИЕ ПРИ УПРАВЛЕНИИ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

*О.В. Тиханьчев, к.т.н., старший научный сотрудник, tow65@yandex.ru
(27 Центральный научно-исследовательский институт Минобороны России,
1-й Хорошевский пр-д, 5, г. Москва, 123007, Россия)*

Условие адекватности управления сложными системами – наличие сведений об их текущем состоянии и условиях функционирования. Обычно такие данные получаются от систем мониторинга обстановки. Но для динамических систем обычный мониторинг не всегда обеспечивает эффективность управления. В ряде случаев ситуацию спасает введение в контур управления обратной связи. Однако этот подход не всегда срабатывает, особенно при управлении крупными распределенными системами, обладающими высокой инерционностью. Для обеспечения эффективности управления необходимо использовать обратную связь: не просто отслеживать состояние системы и окружающей среды, но и получать информацию об их возможных изменениях заблаговременно, то есть использовать методы прогнозирования.

В настоящее время принято разделять все методы прогнозирования на активные, оценивающие возможные последствия принимаемых решений, и пассивные, обеспечивающие прогноз изменения состояния в текущих условиях.

Предлагается использовать пассивное прогнозирование для формирования обратной связи активного типа, позволяющей формировать управляющие воздействия заблаговременно с учетом прогноза развития обстановки.

Анализ состава исходных данных и требований по оперативности прогноза позволяет сделать вывод о целесообразности применения для организации активного мониторинга в существующих условиях именно моделей на основе временных рядов.

Ключевые слова: прогнозирование, активный мониторинг, математическое моделирование, управление динамическими системами.

Несмотря на постоянное развитие положений теории управления, до настоящего времени остается актуальной проблема повышения эффективности управления распределенными динамическими системами. Основа решения этой проблемы – комплексное совершенствование организации основных этапов типового цикла управления.

Как показывает практика, важнейшими этапами любого цикла управления являются оценка обстановки и контроль выполнения принимаемых решений. Первый этап служит основой для выработки решения, второй – для завершения цикла управления и перехода к следующей итерации. Условием успешности выполнения указанных этапов является наличие актуальной информации о параметрах состояния управляемой системы, взаимодействующих (конкурирующих) систем и условиях их функционирования.

Методы оценки состояния управляемой системы

Для получения указанных данных в состав автоматизированных систем управления включаются системы сбора данных обстановки и средства ее мониторинга. Мониторинг состояния осуществляется путем периодического считывания установленных оператором параметров управляемой системы, известных (предполагаемых) параметров противодействующих и взаимодействующих систем, а также условий их функционирования с последующим сравнением их с критериальными значениями. Такой подход можно отнести к самой простой и наиболее распространенной форме мо-

нитинга, когда агрегированные текущие параметры периодически сравниваются с критериями и сигнал «не норма» выдается в момент выхода параметра за установленные границы.

Такая организация мониторинга не сложна и, в первую очередь, требует оперативного получения и обработки составляющих контролируемых параметров. Однако в условиях управления распределенными динамическими системами простой мониторинг не всегда обеспечивает адекватность управления. Как показывает практика, в ряде случаев могут наблюдаться запаздывание сигналов управления, несоответствие мощности управляющих воздействий требованиям достижения цели и т.п.

С алгоритмической точки зрения, описанный подход к организации мониторинга аналогичен принципу использования контура обратной связи, реализуемому в системах автоматического управления. Этот принцип надежный, показавший на практике достаточно высокую эффективность. В то же время при управлении сложными распределенными системами, обладающими высокой инерционностью, простая обратная связь эффективна далеко не всегда. В процессе сбора информации и реализации управляющих воздействий система настолько меняет свое состояние, что сигналы обратной связи просто запаздывают. Для решения проблемы необходимо иметь возможность прогнозирования изменений состояния системы на некоторый период времени, обеспечивающий опережение принятия решения, то есть создать некоторую дополненную обратную связь. Основа такой связи – прогнозирование состояния системы на некоторый период времени, опережающий ее текущее

положение с учетом реализации управляющих воздействий.

В специальной литературе иногда принято разделять прогнозирование на активное и пассивное [1, 2]. Первое обеспечивает оценку возможных последствий принимаемых решений, то есть дает ответ на вопрос «что будет, если сделать то-то». Второе оценивает возможное состояние системы через некоторое время, если условия ее функционирования останутся неизменными, то есть отвечает на вопрос «что будет, если ничего не предпринимать». Как показывает практика, для управления динамическими системами с высоким уровнем собственной инерционности целесообразно использовать пассивное прогнозирование, которое в условиях получения прогнозов с высокой периодичностью можно условно назвать активным мониторингом. Формально активный мониторинг – это то же пассивное прогнозирование, но с более жесткими требованиями к оперативности и непрерывности прогноза.

Прогнозирование изменения состояния управляемой системы как одна из основ эффективности управления

Пояснить сущность активного мониторинга позволяет аналогия с использованием принципов обратной связи в *системах автоматического управления* (САУ), к которым с определенными допущениями может быть сведена любая система «управляемый объект–автоматизированная система управления–органы управления». Практика показывает, что охват управляющих звеньев САУ обратной связи улучшает общие характеристики системы, но при его применении могут возникать определенные проблемы. Например, наличие простой или жесткой обратной связи повышает степень астатизма системы, что при определенных условиях приводит к существенному снижению качества управления. Для компенсации влияния этой проблемы в состав обратной связи, как правило, включают дифференцирующие звенья с уровнем дифференциации не ниже первого порядка $W_{oc}(p)$, то есть делают связь гибкой. Именно наличие гибкой обратной связи обеспечивает аналог активного мониторинга, повышая устойчивость системы к случайным возмущениям и систематическим ошибкам (рис. 1).

Указанная система в варианте отрицательной обратной связи может быть описана уравнением

$$V(p) = \frac{W_{пц}(p)}{1 + W_{пц}(p)W_{oc}(p)}.$$

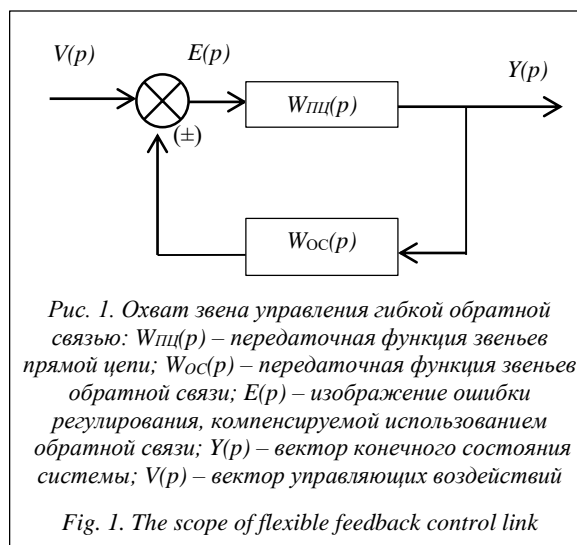
Повышение передаточных свойств такой системы, улучшение ее фазочастотной характеристики в САУ будут соответствовать повышению оперативности принятия решений при сохранении уровня их обоснованности в автоматизированных

системах управления. При этом, как показывает анализ уравнения управляемой системы, наибольшая управляемость достигается, когда передаточная функция обратного звена соизмерима с передаточной функцией самой системы. Существенное отклонение величин передаточных функций либо сводит на нет положительное влияние обратной связи, либо вызывает рассогласование управляющего и выходного сигналов. В свою очередь, параметры обратной связи и определяются характеристиками используемых в системе управления средств прогнозирования, то есть компонентов активного мониторинга.

Активный мониторинг в роли гибкой обратной связи, хотя и похож по принципу организации на пассивное прогнозирование, имеет ряд отличий от него по оперативности получения и обработки исходных данных, а следовательно, и по применяемым для этого математическим методам. На основании этого он может быть обоснованно выделен в отдельный класс задач. Основные требования к активному мониторингу: получение прогнозных оценок, сбалансированных по показателям оперативности между требованиями обеспечения динамического управления системой и минимально достаточной для принятия решений точности прогноза. Эти требования определяют выбор методов прогнозирования, которые могут быть использованы в данном процессе.

Методы прогнозирования состояния процессов и систем

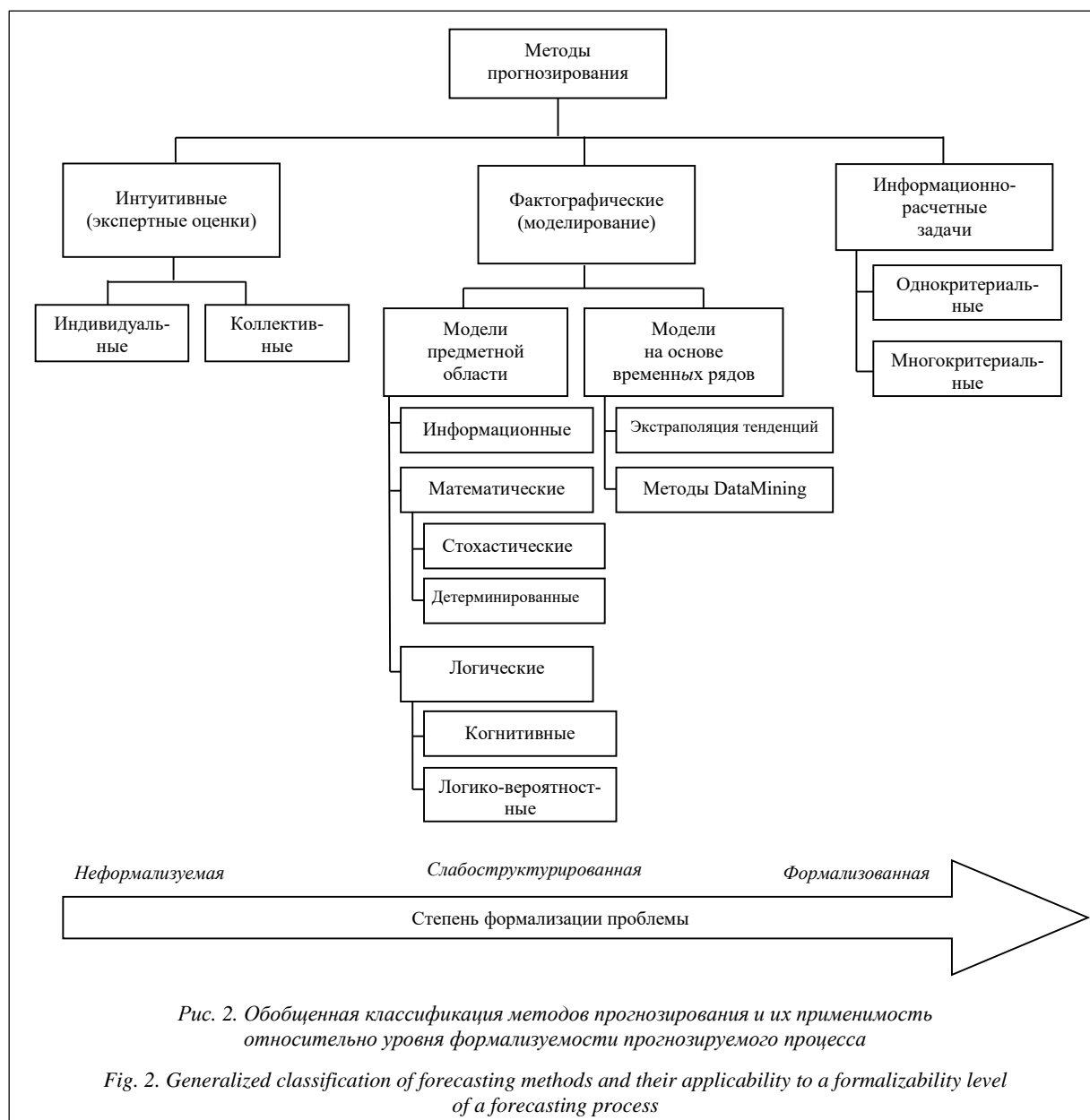
В настоящее время на практике используется довольно широкий спектр методов прогнозирования. Среди них принято выделять две основные группы методов получения прогнозных оценок: интуитивные, имеющие дело с субъективными суждениями, и формальные, использующие расчетные методики и математические модели. Эти



методы реализуются через применение различного математического аппарата: от экспертных оценок для интуитивных методов до сложных математических моделей, реализуемых в фактографических подходах [3–5]. Применяемый в той или иной ситуации математический аппарат прогнозирования зависит от сложности системы и степени формализуемости проблемы (рис. 2). Классификация, приведенная на рисунке, авторская, и с ней можно спорить, но общую тенденцию состояния методов прогнозирования она отражает. При абсолютной формализации предметной области в ряде случаев для получения прогнозных параметров вектора состояния системы могут использоваться даже простые расчетные задачи, но это, скорее, исключение, чем правило. Чаще всего разработчику приходится использовать средства для учета экспертных оценок или математическое моделирование.

Задача активного мониторинга, как показывает анализ содержания данной проблемы, может решаться исключительно с применением аппарата математического моделирования. Для реализации процесса активного мониторинга могут быть использованы преимущественно фактографические типы моделей [1], которые подразделяются на модели временных рядов и модели предметной области [6, 7].

В данной постановке модели временных рядов – это математические модели прогнозирования, которые обеспечивают поиск зависимости будущих значений параметров от истории развития самого процесса и построение по этой зависимости прогноза [8, 9]. Эти модели универсальны для различных предметных областей, то есть их общий вид не меняется в зависимости от природы временного ряда.



Модели предметной области – такие математические модели прогнозирования, для построения которых используют законы предметной области. В этих моделях используются математические, логические и вероятностные зависимости, свойственные конкретной предметной области. Такого рода моделям присущи индивидуальный подход к разработке математического аппарата и детальный учет всех значимых параметров системы и среды ее функционирования.

Каждая из описанных групп моделей обладает набором свойств, определяющим их пригодность или непригодность к применению в тех или иных областях. Задача активного мониторинга выдвигает свои требования к используемым моделям, в основе которых оперативность получения прогноза и наличие в составе исходной информации больших массивов динамично меняющихся данных. В таких условиях применение моделей предметной области, большинство из которых требуют временных затрат на моделирование и ввода большого количества исходных данных, даже несмотря на их высокую точность, проблематично. В то же время модели временных рядов, обладающие несколько меньшей точностью, но менее требовательные к вводу исходных данных и более оперативные, в большинстве случаев оказываются предпочтительнее. Хотя их достоинства, как это часто бывает, являются продолжением их недостатков. Прежде всего к таким недостаткам относится слабый учет изменения внешних условий функционирования управляемой системы и других факторов, не описанных статистикой предыдущих состояний. В перспективе с развитием методов автоматического сбора и ввода данных модели предметной области смогут достойно участвовать в разрешении перманентного противоречия между точностью и оперативностью прогноза, в том числе при использовании в автоматизированных системах поддержки принятия решений [10–12]. В настоящее время основная область их применения гносеологическая, а в сфере управления – прогнозирование последствий принимаемых решений при планировании в ситуациях, не критичных к длительности цикла управления.

Выводы

Таким образом, анализ состава исходных данных и требований по оперативности прогноза позволяет сделать вывод о целесообразности применения для организации активного мониторинга в существующих условиях именно моделей на основе временных рядов.

К математическим методам, реализующим подобные модели, можно отнести

– методы экстраполяции тенденций (параболическая экстраполяция, метод подбора функций, модель прогнозирования авторегрессии произинте-

грированного скользящего среднего с учетом внешнего фактора и другие);

– методы обработки данных DataMining (методики автоматического поиска типовых сценариев SmartData, методы построения системы рассуждений на аналогиях case based reasoning – CBR и т.п.).

Отметим, что предметная область активного мониторинга до настоящего времени была недостаточно развита и слабо реализована в практике управления. Данная ситуация обусловлена как организационными, так и технологическими причинами. Организационные причины в том, что долгое время просто не было существенной необходимости в высокой оперативности управления: большинство управляемых распределенных систем были инертными и позволяли обойтись методами обычного мониторинга. Технологические причины порождались недостаточными возможностями применяемых в практике управления информационных технологий.

Сейчас ситуация изменилась. С повышением динамичности управляемых систем, развитием скоростных компьютерных сетей и методов работы с большими массивами информации Big Data [1] подобные задачи обязательно будут востребованы в составе ПО автоматизированных систем управления. Поэтому необходимо обеспечить развитие теоретических положений и практических методов реализации технологий активного мониторинга состояния распределенных систем. А для запуска этого процесса нужно прежде всего признать наличие указанной проблемы и классифицировать ее.

Литература

1. Тиханьчев О.В., Саяпин О.В. Оперативное прогнозирование развития обстановки как основа успешного управления применением войск (сил) // Военная мысль. 2015. № 4. С. 3–7.
2. Новиков Д.А., Чхартишвили А.Г. Активный прогноз. М.: Изд-во ИПУ РАН, 2002. 101 с.
3. Цигичко В.Н. Модели в системе принятия военно-стратегических решений в СССР. М.: Империя-Пресс, 2005. 96 с.
4. Тиханьчев О.В. Общие подходы к обеспечению автоматизированной поддержки принятия решений. М.: Эдитус, 2014. 64 с.
5. Ляпин В.Р., Барвиненко В.В. Единая информационно-моделирующая среда в системах военного назначения // Военная мысль. 2015. № 4. С. 72–78.
6. Выпасняк В.И., Гуральник А.М., Тиханьчев О.В. Моделирование военных действий – история, состояние, перспективы развития // Военная мысль. 2014. № 7. С. 28–37.
7. Денисов В.Н., Саяпин О.В., Тиханьчев О.В. О месте математического моделирования в работе органов военного управления // Военная мысль. 2016. № 5. С. 28–33.
8. Дружинина Н.Г., Трофимова О.Г., Трофимов С.П. Информационно-коммуникационная система управления движением и анализа эффективности работы городского электротранспорта // Программные продукты, системы и алгоритмы. 2013. № 2. URL: <http://swsys-web.ru/information-and-communication-systems.html> (дата обращения: 24.08.16).
9. Болодурина И.П., Решетников В.Н., Таспаева М.Г. Применение и адаптация нейросетевых технологий в задаче идентификации динамических объектов // Программные продукты, системы и алгоритмы. 2013. № 2. URL: <http://swsys-web.ru/>

application-of-neural-network-technology.html (дата обращения: 24.08.16).

10. Виноградов Г.П., Семенов Н.А. Модели прогнозирования в интеллектуальных системах // Программные продукты и системы. 2007. № 4. С. 80–82.

11. Решетников А.Г., Керимов Т.А., Ульянов С.В. Робастное интеллектуальное управление на основе технологии кванто-

вых вычислений // Программные продукты и системы. 2013. № 4. С. 192–199.

12. Башлыков А.А., Еремеев А.П. Семантические системы реального времени для интеллектуальной поддержки принятия решений при управлении сложными технологическими объектами // Информационно-измерительные и управляющие системы. 2013. № 5. С. 49–57.

Software & Systems

DOI: 10.15827/0236-235X.117.040-044

Received 25.08.16

2017, vol. 30, no. 1, pp. 40–44

FORECASTING IN DYNAMIC SYSTEM CONTROL

O.V. Tikhanychev¹, Ph.D. (Engineering), Senior Researcher, tow65@yandex.ru

¹ 27 Central Research Institute of the Ministry of Defense of Russia, 1st Khoroshevsky lane 5, Moscow, 123007, Russian Federation

Abstract. The condition of complex systems management adequacy is a stock of information on their current state and operation conditions. Typically, such data is obtained from environment monitoring systems. But usual monitoring of dynamic systems does not always provide effective management. In some cases it is appropriate to introduce a feedback to a control loop. But this approach does not always work, especially when managing large distributed systems with high inertia. In order to ensure management efficiency it is necessary to use feedback. It means not only monitor the status of a system and the environment, but also to obtain information on their possible changes in advance, that is to use forecasting methods.

Now it is common practice to divide all forecasting methods on active (which evaluate possible consequences of the decisions made) and passive (which provide a forecast of changing a state under current conditions). It is proposed to use passive forecasting for active type feedback formation. It allows a user to form control actions in advance taking into account the forecast of a situation development

Keywords: forecasting, proactive monitoring, mathematical modeling, dynamic systems control.

References

1. Tikhanychev O.V., Sayapin O.V. Development operational forecasting as a base of successful control of military forces handling. *Voennaya mysl* [Military Thought]. 2015, no. 4, pp. 3–7 (in Russ.).
2. Novikov D.A., Chkhartishvili A.G. *Aktivny prognoz* [Active Forecasting]. Moscow, V.A. Trapeznikov Institute of Control Sciences of RAS Publ., 2002, 101 p.
3. Tsigichko V.N. *Modeli v sisteme prinyatiya voenno-strategicheskikh resheniy v SSSR* [Models in a military strategic decision-making system in the USSR]. Moscow, Imperium-Press, 2005, 96 p.
4. Tikhanychev O.V. *Obshchie podkhody k obespecheniyu avtomatizirovannoy podderzhki prinyatiya resheniy* [General approaches to automated decision-making support]. Moscow, Editus, 2014, 64 p.
5. Lyapin V.R., Barvinenko V.V. Common information modeling environment in military systems. *Voennaya mysl* [Military Thought]. 2015, no. 4, pp. 72–78 (in Russ.).
6. Vypasnyak V.I., Guralnik A.M., Tikhanychev O.V. Military operation simulation. History, state, development prospects. *Voennaya mysl* [Military Thought]. 2014, no. 7, pp. 28–37 (in Russ.).
7. Denisov V.N., Sayapin O.V., Tikhanychev O.V. On the place of math modeling in operation of command and control organizations. *Voennaya mysl* [Military Thought]. 2016, no. 5, pp. 28–33 (in Russ.).
8. Druzhinina N.G., Trofimova O.G., Trofimov S.P. Information and communication system to control movement and analyze performance of urban electric transport. *Programmnye produkty, sistemy i algoritmy* [Software Products, Systems and Algorithms]. 2013, no. 2. Available at: <http://swsys-web.ru/information-and-communication-systems.html> (accessed August 24, 2016).
9. Bolodurina I.P., Reshetnikov V.N., Taspava M.G. Application and adaptation of neural network technologies in the problem of dynamic object identification. *Programmnye produkty, sistemy i algoritmy* [Software Products, Systems and Algorithms]. 2013, no. 2. Available at: <http://swsys-web.ru/application-of-neural-network-technology.html> (accessed August 24, 2016).
10. Vinogradov G.P., Semenov N.A. Forecasting models in intelligent systems. *Programmnye produkty i sistemy* [Software & Systems]. 2007, no. 4, pp. 80–82 (in Russ.).
11. Reshetnikov A.G., Kerimov T.A., Ulyanov S.V. Robust intelligent control based on the quantum computing technology. *Programmnye produkty i sistemy* [Software & Systems]. 2013, no. 4, pp. 192–199 (in Russ.).
12. Bashlykov A.A., Eremeev A.P. Semiotics systems of real time for intelligent decision support at management of complex technological objects. *Informatsionno-izmeritelnye i upravlyaushie sistemy* [Information-measuring and Control Systems]. 2013, no. 5, pp. 49–57 (in Russ.).

УДК 681.516.5

DOI: 10.15827/0236-235X.339.045-050

Дата подачи статьи: 1.11.16

2017. Т. 30. № 1. С. 45–50

СИСТЕМЫ АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ ОБЪЕКТАМИ С ЗАПАЗДЫВАНИЕМ: РОБАСТНОСТЬ, БЫСТРОДЕЙСТВИЕ, СИНТЕЗ

*В.З. Тхан, аспирант, dungvietthan@gmail.com;**Д.Ю. Берчук, аспирант, berchukdy@gmail.com**(Национальный исследовательский Томский политехнический университет,
просп. Ленина, 30, г. Томск, 634050, Россия)*

Рассматривается задача синтеза систем автоматического управления объектами с запаздыванием. Особенность ее постановки связана с формированием и решением уравнения синтеза численным методом без аппроксимации передаточной функции звена запаздывания. Этим создаются потенциальные возможности для повышения точности синтеза регуляторов. В работе они раскрываются за счет привлечения численного подхода к синтезу регуляторов – вещественного интерполяционного метода. Метод позволяет найти стабилизирующий регулятор по модели желаемой системы автоматического управления, известному неаппроксимированному математическому описанию объекта управления и принятой структуре регулятора. Вторая особенность задачи также связана с принятым методом и звеном запаздывания и реализуется на этапе получения желаемой модели.

Для усиления демонстрационных возможностей и наглядности подхода в работе использован минимальный набор показателей – перерегулирование и время установления переходного процесса. Модель желаемой системы совместно с неаппроксимированным описанием объекта управления и принятой структурой регулятора позволили сформировать более точное уравнение синтеза по сравнению с традиционным подходом, основанным на приближении модели звена запаздывания дробно-рациональным выражением.

Рассмотрены численный способ решения уравнения синтеза, а также исследование процедуры синтеза и свойств робастности систем с запаздыванием. На этой основе предложен алгоритм достижения максимального быстродействия систем автоматического управления с запаздыванием в условиях поддержания перерегулирования в заданных пределах и показан механизм ухудшения робастности синтезированной системы при увеличении запаздывания сигнала в объекте управления даже в условиях подстройки параметров регулятора под изменяющееся время запаздывания.

Результаты проиллюстрированы расчетными примерами.

Ключевые слова: синтез систем управления с запаздыванием, робастность, быстродействие.

Системы автоматического управления (САУ) объектами с запаздыванием составляют обширный и важный для практики класс динамических систем. Расчет таких систем имеет специфические особенности в силу двух обстоятельств. Во-первых, это физические особенности, когда реакция объекта задерживается на время τ относительно момента подачи сигнала на вход объекта. Во-вторых, это же явление вносит свои особенности в расчетные процедуры САУ из-за наличия в их математическом описании трансцендентной составляющей $e^{-p\tau}$, отражающей существование запаздывания τ . С позиций синтеза систем с запаздыванием обе особенности учитываются при построении регуляторов на основе популярного предиктора Смита [1]. Подход эффективен, но имеет свои ограничения [2, 3]. В частности, в условиях переменного запаздывания система управления может значительно изменять свои свойства вплоть до потери устойчивости. К этому можно добавить, что параметры дробно-рациональной части объекта практически всегда также изменяются с течением времени, что ухудшает качество управления. В настоящей работе делается попытка поиска средств, позволяющих раскрывать робастные свойства САУ с запаздыванием на этапе синтеза регулятора и представлять их в наглядной форме. Аналитическое решение такой задачи, очевидно, не-

возможно из-за трудностей действий с функцией $e^{-p\tau}$, входящей в модель САУ. По этой причине в работе используется один из численных методов, позволяющих оперировать описаниями САУ с запаздыванием.

Выбор численного метода расчета систем с запаздыванием

При решении задач автоматического управления основным математическим аппаратом является преобразование Лапласа, которое позволяет заменить операции в области времени более простыми действиями в области изображений. Получение таких изображений осуществляется по формуле прямого преобразования

$$F(p) = \int_0^{\infty} f(t)e^{-pt} dt, p = \delta + j\omega. \quad (1)$$

Она удобна для аналитических преобразований, однако переход к численным действиям в общем случае затруднителен из-за вычислительных сложностей. Поэтому привлекают частные случаи преобразования (1), вытекающие из двух возможных трансформаций комплексной переменной $p = \delta + j\omega$. Первый вариант определяется условием $\delta = 0$, при котором комплексная переменная вырождается в чисто мнимую $p = j\omega$. Получаемые изображения $F(j\omega)$ соответствуют частотному методу на основе

преобразования Фурье, они легко преобразуются в численные модели вида $F(j\omega_i)$, $i=1, 2, \dots$ [4]. Вторым вариантом вытекает из преобразования (1) при условии $\omega=0$, когда комплексная переменная вырождается в вещественную $p=\delta$ [5, 6]. Получаемые функции-изображения

$$F(\delta) = \int_0^\infty f(t)e^{-\delta t} dt, \delta \in [C, \infty], C \geq 0, \quad (2)$$

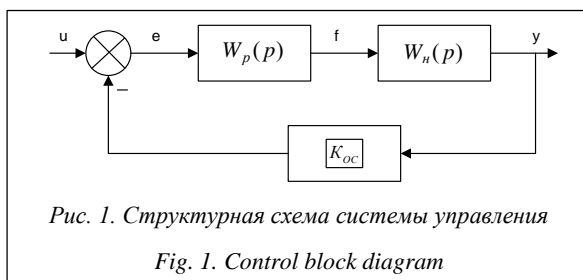
являются вещественными функциями вещественной переменной, что важно с позиций объема вычислений и практических задач управления. Функции $F(\delta)$ позволяют достаточно просто перейти к дискретным представлениям $F(\delta_i)$, $i=1, 2, \dots, \eta$, привлечь для действий над ними хорошо разработанные численные методы. В отношении САУ с запаздыванием, которые являются предметом рассмотрения в настоящей работе, отметим: модели вида $F(\delta_i)$, $i=1, 2, \dots, \eta$, встречаются существенно меньше трудности при их преобразовании по сравнению с аналитическим подходом, так как действия осуществляются в вещественной арифметике – над числами $e^{-\delta_i \tau}$, $i=1, 2, \dots, \eta$.

В формуле вещественного преобразования (2) предусмотрено ограничение переменной δ снизу величиной $C \geq 0$, что обеспечивает сходимость интеграла для широкого класса функций-оригиналов $f(t)$, охватывающих практически все важные виды сигналов динамических систем.

Синтез регуляторов САУ с запаздыванием численным методом

Выбранный численный метод позволяет находить регуляторы автоматических систем, управляющих так называемыми особыми объектами, в том числе с запаздыванием, с распределенными параметрами [5, 6]. Воспользуемся такой возможностью для рассмотрения быстродействия САУ с запаздыванием и их робастных свойств. С этой целью рассмотрим конкретные задачи, в рамках которых можно выполнять численные расчеты, анализировать результаты и делать обобщающие выводы.

В качестве примера управления использовались экструдеры, предназначенные для изготовления пластиковой нити для 3D-принтеров [7], а также близкие к ним объекты [8], управление которыми сводится к типовой схеме, представленной на рисунке 1.



Система содержит регулятор, объект управления и цепь обратной связи. Для объекта [8] элементы САУ представлены описаниями

$$W_n(p) = \frac{0,5}{4p^2 + 8p + 1} \cdot e^{-3p},$$

$$W_p(p) = \frac{b_1 p + b_0}{p(a_1 p + 1)}, k_{oc} = 1.$$

По технологии требуется поддерживать перерегулирование σ в заданных пределах:

$$\sigma_3 - \Delta\sigma \leq \sigma \leq \sigma_3 + \Delta\sigma. \quad (3)$$

Обозначения σ_3 , $\Delta\sigma$, σ в (3) соответствуют заданному значению перерегулирования $\sigma_3 = 5\%$, достигнутой величине σ и допустимому отклонению $\Delta\sigma = |\sigma_3 - \sigma|$, $\Delta\sigma = \pm 1\%$. Еще одно требование – время установления переходного процесса должно быть минимальным:

$$t_y \rightarrow \min. \quad (4)$$

В этих условиях требуется синтезировать регулятор, не прибегая к аппроксимации звена запаздывания, рассмотреть робастные свойства системы, а также выяснить зависимость этих свойств от величины запаздывания.

Для решения задачи составим уравнение синтеза $W_{*}(p) \cong \frac{W_p(p)W_o(p)}{1+W_p(p)W_o(p)k_{oc}}$. Входящая в левую

часть уравнения желаемая передаточная функция $W_{*}^3(p)$ найдется по исходным данным. Неопределенность возникает лишь при задании желаемого времени установления t_y^{*} , приближенным ориентиром для которого являются коэффициенты знаменателя модели объекта. В условиях численного метода и итерационного приближения к искомому решению неточность такого рода не имеет существенного значения. Более того, как будет ясно из дальнейшего, эта неоднозначность и итерационная процедура позволяют получить множество решений, из которых можно выбрать наиболее подходящее, привлекая дополнительные критерии. Желаемое время переходного процесса t_y^{*} будем рассматривать в качестве переменного параметра, обеспечивая его минимизацию при выполнении других условий. Для начальных расчетов оно было принято равным 115 с, что позволило получить методом [5] желаемую передаточную функцию:

$$W_{*}^3(p) = \frac{21p + 1}{794p^2 + 42p + 1}.$$

В правой части уравнения синтеза содержится модель объекта управления, которая известна, и передаточная функция регулятора, вид которой выбран на основе априорных данных:

$$W_p(p) = \frac{b_1 p + b_0}{p(a_1 p + 1)}.$$

Для упрощения уравнения синтеза, сопоставляющего желаемую и синтезируемую системы в за-

мкнутом состоянии, перейдем к равносильному уравнению $W_p^p(p) \approx W_p(p) \cdot W_n(p)$, описывающему эти же системы в разомкнутом состоянии. В левой части уравнение содержит передаточную функцию разомкнутой системы $W_p^p(p) = \frac{p + 0,048}{p(37,81p + 1)}$, которая находится однозначно по уже известной функции $W_n^3(p)$.

В результате получаем расчетное соотношение изображений по Лапласу:

$$\begin{aligned} \frac{p + 0,048}{p(37,81p + 1)} &\cong \\ &\cong \frac{b_1 p + b_0}{p(a_1 p + 1)} \cdot \frac{0,5}{4p^2 + 8p + 1} \cdot e^{-3p}. \end{aligned} \quad (5)$$

Дальнейшие шаги могут быть сделаны в рамках известных инженерных методов. Они связаны с заменой передаточной функции звена запаздывания на дробно-рациональное выражение, адаптированное к классическим методам. Другой вариант – использование принятого численного метода. В последнем случае появляется возможность исключения аппроксимации звена запаздывания, что позволяет надеяться на повышение точности решения уравнения. Для его реализации перейдем в уравнении (5) к вещественным изображениям. Такой переход для устойчивых звеньев и систем можно сделать формально заменой переменных: $p \rightarrow \delta$ [6]. Очевидно, что все полюсы функций в соотношении (5) расположены в левой полуплоскости, так что можно воспользоваться формальным приемом, что приводит к форме

$$\begin{aligned} \frac{\delta + 0,048}{\delta(37,81\delta + 1)} &\cong \\ &\cong \frac{b_1 \delta + b_0}{\delta(a_1 \delta + 1)} \cdot \frac{0,5}{4\delta^2 + 8\delta + 1} \cdot e^{-3\delta}. \end{aligned} \quad (6)$$

Для численного решения уравнения (6) заменим входящие в него функции их численными представлениями вида $F(\delta_i)$, $i=1, 2, 3$, в соответствии с рекомендациями [6]. В результате получим систему линейных алгебраических уравнений (СЛАУ):

$$\begin{aligned} \frac{\delta_i + 0,048}{\delta_i(37,81\delta_i + 1)} &= \\ &\cong \frac{b_1 \delta_i + b_0}{\delta_i(a_1 \delta_i + 1)} \cdot \frac{0,5}{4\delta_i^2 + 8\delta_i + 1} \cdot e^{-3\delta_i}, i = 1, 2, 3. \end{aligned} \quad (7)$$

При формировании СЛАУ для упрощения задачи узлы интерполирования приняты равномерными: $\delta_i = i\delta_1$, $i=1, 2, 3$. В этом случае первый узел δ_1 можно использовать в качестве инструментальной переменной, что позволяет из множества решений выбрать такое, которое в наибольшей степени отвечает предъявляемым требованиям, в частности по критерию (3).

По итогам принятых уточнений констатируем: имеются заданное значение перерегулирования σ_3 , условие (4) и два переменных параметра – t_y^* и δ_1 , с помощью которых необходимо достичь заданных показателей качества САУ. Для выстраивания приоритетов будем считать перерегулирование ведущим параметром, учитывая его интегральный смысл, так как он отражает не только качество работы САУ, но и косвенно запас устойчивости. Такой подход известен и успешно используется [9]. Для объединения трех выделенных параметров воспользуемся формой $\sigma=f(\delta_1)$ для различных t_y^* . Это позволит оптимизировать быстродействие синтезируемых систем, оценить их робастность [10], а также влияние величины запаздывания объекта управления на эти свойства САУ.

Быстродействие САУ с запаздыванием

Быстродействие создаваемой системы при прочих равных условиях будет зависеть от желаемого времени установления переходного процесса, задаваемого при формировании эталонной передаточной функции. На рисунке 2 приведены результаты синтеза регуляторов для нескольких значений t_y^* . Итоги расчетов даны в виде графиков $\sigma=f(\delta_1)$ для нескольких значений желаемого времени установления: $t_{y1}^* = 100$ с, $t_{y2}^* = 115$ с, $t_{y3}^* = 130$ с. Видно, что положение кривой $\sigma=f(\delta_1)$ зависит от задаваемой величины t_y^* . Изменяя ее, можно найти такое положение графика $\sigma=f(\delta_1)$, при котором перерегулирование σ соответствует заданному значению σ_3 с заранее определенной точностью. Таким образом, на этапе синтеза регулятора можно обеспечить выполнение условий (1) и (2). Для принятого расчетного варианта, как показывает рисунок 2, приемлемой оказывается итерация, соответствующая времени $t_{y2}^* = 115$ с.

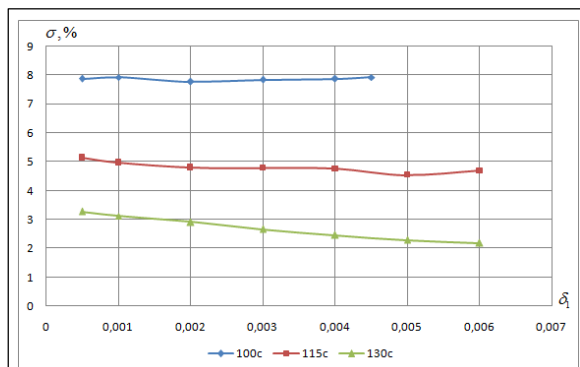


Рис. 2. Графики зависимости $\sigma=f(\delta_1)$ для различных значений t_y^*

Fig. 2. The graphs $\sigma=f(\delta_1)$ for values t_y^*

Обратим внимание на важную для практики возможность синтезировать регуляторы, которые ограничивают перерегулирование не только сверху, как это обычно бывает, но и снизу. Подчеркнем, что речь идет не о простейших САУ, а о системах с запаздыванием.

Робастность САУ с запаздыванием

Робастное свойство САУ многогранно, поэтому обычно ограничиваются какими-либо конкретными показателями системы. В данном случае за такой показатель вновь примем перерегулирование в силу его обобщающих свойств по отношению к показателям устойчивости и качества САУ. В [10] показано, что при определенных условиях изменение величины δ_1 приближенно соответствует изменению значений физических параметров САУ и, следовательно, коэффициентов ее передаточной функции. Это позволяет оценивать робастность системы по функции $\sigma = f(\delta_1)$ – величине той ее части, которая удовлетворяет условию (3). Для расчетов и сравнительного анализа вариантов используют количественную оценку робастности, которая определяется максимальной величиной интервала $[\delta_{1\min}, \delta_{1\max}]$, внутри которого выполняется условие (3). Обозначим такую оценку как

$$r_\sigma = \delta_{2\min} - \delta_{1\max}, \quad (8)$$

указывая индексом σ на отношение показателя робастности к перерегулированию.

Воспользуемся (8) для исследования влияния величины запаздывания на робастные свойства САУ. С этой целью сохраним неизменной дробно-рациональную часть модели объекта управления и структуру регулятора, а величину запаздывания будем изменять, оценивая робастность синтезируемых САУ по величине r_σ . Для оценивания итогов вновь обратимся к наглядной зависимости $\sigma = f(\delta_1)$. Результаты расчетов для нескольких величин запаздывания представлены на рисунке 3.

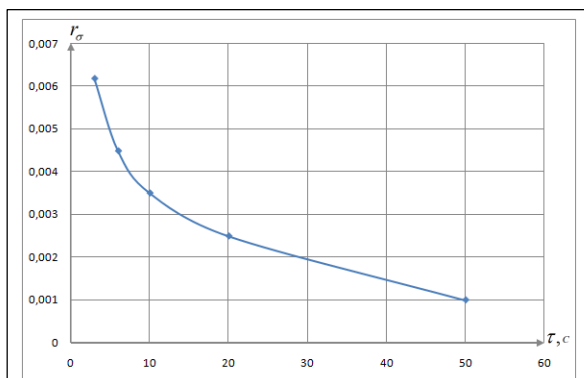


Рис. 3. График зависимости показателя робастности r_σ от величины запаздывания

Fig. 3. A graph of dependence of a robustness indicator on a delay value

График показывает, что увеличение запаздывания существенно ухудшает робастные свойства САУ по перерегулированию. При этом подчеркнем, что график получен в условиях настройки регулятора САУ на каждое новое изменение времени запаздывания. В тех случаях, когда такая подстройка не выполняется, робастность снижается намного сильнее.

Есть основания предполагать, что показатели качества и устойчивости имеют примерно такую же зависимость от величины запаздывания. Представленные сведения свидетельствуют о том, что в рамках рассматриваемого численного подхода имеется возможность оценивать влияние изменения запаздывания на свойства САУ, в частности на перерегулирование. С этих позиций можно дать практическую рекомендацию по выбору величины переменной δ_1 : она должна находиться в середине интервала $[\delta_{1\min}, \delta_{1\max}]$. Возможности расширения робастных свойств САУ увеличиваются, если перерегулирование ограничивается только сверху, что имеет место во многих практических задачах. При этом достаточно сместить рабочую точку δ_1 в сторону начала интервала $[\delta_{1\min}, \delta_{1\max}]$.

Повышение точности безаппроксимационного решения уравнения синтеза САУ с запаздыванием

Использование неаппроксимированного уравнения синтеза САУ с запаздыванием для повышения точности таких систем привлекательно и в каких-то случаях может оказаться эффективным. Проверим эту возможность, используя прежний объект управления. Для сравнения с традиционным подходом, основанным на приближении звена запаздывания рациональной дробью, выберем два достаточно сильно различающихся по точности приближения звена запаздывания $e^{-\tau p}$:

$$W_{np1}(p) = \frac{b_0}{a_1 p + 1}, \quad (9)$$

$$W_{np2}(p) = \frac{b_1 p + b_0}{a_2 p^2 + a_1 p + 1}. \quad (10)$$

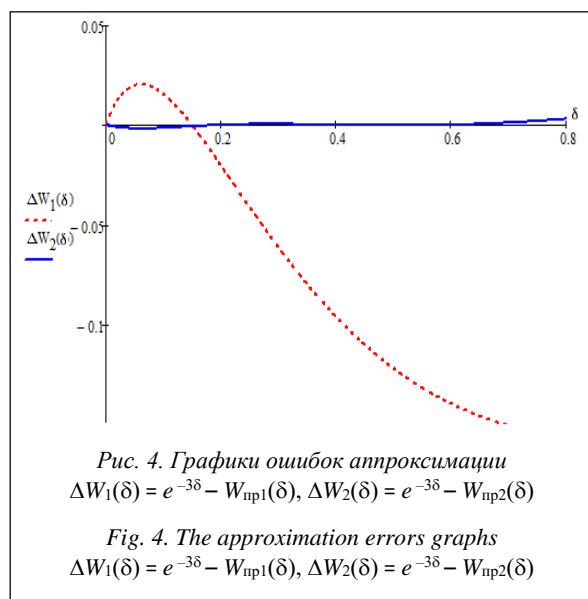
Вычислим неизвестные коэффициенты численным методом [6]. Для структур (9) и (10) получены две модели объекта управления:

$$W_{H1}(\delta) = \frac{0,5}{4\delta^2 + 8\delta + 1} \cdot \frac{1}{4,865p + 1},$$

$$W_{H2}(\delta) = \frac{0,5}{4\delta^2 + 8\delta + 1} \cdot \frac{-0,752p + 1}{2,816p^2 + 2,177p + 1}.$$

Для получения представления о точности приближения воспользуемся наглядным представлением ошибки в области вещественных изображений как функции $\Delta W(\delta) = e^{-\rho\tau} - W_{np}(\delta)$, где $W_{np}(\delta)$ – аппроксимированная вещественная передаточная

функция звена запаздывания. Результаты аппроксимации представлены на рисунке 4.



Сопоставляя графики, можно констатировать, что погрешность аппроксимации в случае функции второго порядка $W_2(p)$ значительно меньше по сравнению с максимально простой функцией первого порядка $W_1(p)$. Об этом можно судить не только по графикам, но и по количественным оценкам, например, вида $\Delta W = \max |\Delta W(\delta)|$.

Таким образом, имеются три модели объекта управления: точная и две приближенные, у одной из которых сравнительно высокая точность. Наличие трех различных по точности моделей позволит установить влияние погрешности представления модели на результаты синтеза САУ. Для оценивания результатов были получены функции $\sigma=f(\delta_1)$, соответствующие принятым моделям объектов. Они отвечают требованию по перерегулированию (3), которое является обязательным требованием синтеза САУ. Выяснилось, что вопреки ожиданиям робастность по перерегулированию для всех трех случаев примерно одинакова. Заметно худшие результаты получены для простейшей модели (8) в том смысле, что робастность оказалась ошибочно завышенной за счет увеличения верхней границы δ_{max} .

В то же время быстродействие САУ оказалось существенно зависимым от точности модели объекта. С понижением точности модели время установления t_y синтезированной САУ увеличивается. Для принятых моделей оно составляет соответственно 149, 151 и 180 с. Это время отсчитывалось для среднего значения инструментальной переменной $\delta_1 = (\delta_{1\text{max}} - \delta_{1\text{min}})/2$, когда робастность каждой системы максимальна. Отметим, что при снижении требований по робастности можно уменьшить время установления t_y , опираясь на функцию $\sigma=f(\delta_1)$.

Заключение

В работе рассмотрены вопросы синтеза регуляторов САУ объектами, имеющими запаздывание, направленные на повышение робастных свойств этих систем. Для проведения исследований был выбран численный путь, позволяющий снизить трудности, связанные с трансцендентной формой модели звена запаздывания. Этот выбор позволил

- составить и решить уравнение синтеза, в котором модель звена запаздывания не требует аппроксимации, что потенциально обеспечивает повышенную точность синтеза САУ;
- использовать возможность синтеза САУ, в которых перерегулирование находится в заданных пределах, а быстродействие близко к максимальному;
- установить наглядную зависимость, отражающую ухудшение робастности САУ при увеличении запаздывания.

Подход может быть реализован не только в рамках синтеза САУ с запаздыванием, но и при построении самонастраивающихся САУ, в которых изменяется как запаздывание, так и другие параметры объекта управления. Для систем этого класса имеются положительные предпосылки [12].

Литература

1. Смит О.Дж.М. Автоматическое регулирование; [пер. с англ. под ред. Е.П. Попова]. М.: Физматгиз, 1963. 847 с.
2. Филимонов А.С. Спектральная декомпозиция динамических систем с запаздываниями: Теория и применения: дисс. д. техн. наук. М.: РУДН, 2003. 426 с.
3. Громов Ю.Ю., Земской Н.А., Лагутин А.В., Иванова О.Г., Тютюнник В.М. Системы автоматического управления с запаздыванием. Тамбов: Изд-во ТГТУ, 2007. 76 с.
4. Козлов О.С., Скворцов Л.М. Решение задач теории автоматического управления в программном комплексе MBTU // Мехатроника, автоматика, управление. 2016. Т. 17. № 1. С. 3–10.
5. Анализ и оптимальный синтез на ЭВМ систем управления; [под ред. А.А. Воронова и И.А. Огурка]. М.: Наука, 1984. 344 с.
6. Гончаров В.И. Вещественный интерполяционный метод синтеза систем автоматического управления. Томск: Изд-во ТПУ, 1995. 100 с.
7. Тхан Вьет Зунг, Берчук Д.Ю. Экспериментальное определение передаточной функции нагревателя экструдера // Молодежь и современные информационные технологии: сб. тр. XIII Междунар. науч.-практич. конф. Томск: Изд-во ТПУ, 2016. С. 261–262.
8. Михалевич С.С., Байдали С.А., Чурсин Ю.А. Графический метод расчета параметров настройки регуляторов, приводящих систему к заданным показателям устойчивости // Промышленные АСУ и контроллеры. 2013. № 10. С. 35–39.
9. Гончаров В.И., Лиспиньш А.В., Рудницкий В.А. Синтез робастных регуляторов низкого порядка // Изв. РАН. Теория и системы управления. 2001. № 4. С. 36–43.
10. Крутова И.Н. Формирование алгоритма управления итерационным процессом настройки параметров в системе с упрощенной эталонной моделью // Автоматика и телемеханика. 1998. № 2. С. 72–84.
11. Денисенко В.В. Компьютерное управление технологическим процессом, экспериментом, оборудованием. М.: Горячая линия-Телеком, 2009. 608 с.
12. Goncharov V. Rudnicki V. Real interpolation method in automatic control systems self-adjustment problem // Systems Science, vol. 36, no. 3, 2010, pp. 35–37.

**TIME DELAY AUTOMATIC CONTROL SYSTEMS:
ROBUSTNESS, RESPONSE TIME, SYNTHESIS****V.D. Than**¹, *Postgraduate Student, dungvietthan@gmail.com***D.Yu. Berchuk**¹, *Postgraduate Student, berchukdy@gmail.com*¹ *National Research Tomsk Polytechnic University, Lenin Ave. 30, Tomsk, 634050, Russian Federation*

Abstract. The article considers synthesis of time delay automatic control systems (ACS). A synthesis statement is connected to equation solving by a numerical technique without approximation of a time delay element. This issue has a potential to improve synthesis accuracy of regulators. A real interpolation method (RIM) has number of advantages in synthesis of time delay ACS. RIM allows finding a regulator that compensates time delay using a targeted model, a defined non-approximated mathematical model of object and a regulator structure. Moreover, RIM allows building transfer function with direct quality factors on the stage of developing a targeted model. It's simpler and more useful for a developer.

The article demonstrates potentials of the method using minimal quantity of factors: overshoot and settings time. The targeted model with a defined non-approximated mathematical model and a regulator structure allows building accurate synthesis equation, which is useful for numerical research of synthesis and robust properties of time delay systems. The article proposes an algorithm of a maximal control rate for a time delay system in terms of specified overshoot. It also shows a negative effect of increasing time delay to robustness of the system. A result is graphically illustrated and proved by a number of examples.

Keywords: synthesis of time delay control systems, robustness, performance.

References

1. Smith O.J. *Automatic control* (Russ.ed.: Popov E.P. Moscow, Fizmatgiz Publ., 1963, 847 p.).
2. Filimonov A.S. *Spektralnaya dekompozitsiya dinamicheskikh sistem s zapazdyvaniyami: teoriya i primeneniya* [Spectral Decomposition of dynamical systems with delays: Theory and applications]. Dr.Sc. thesis. Moscow, PFUR Publ., 2003.
3. Gromov Yu.Yu., Zemskoy N.A., Lagutin A.V., Ivanova O.G., Tyutyunnik V.M. *Sistemy avtomaticheskogo upravleniya s zapazdyvaniem* [Automatic control systems with delay]. Tambov, TSTU Publ., 2007, 76 p.
4. Kozlov O.S., Skvortsov L.M. Solving problems in the theory of automatic control software package MVTU. *Mekhatronika, avtomatika, upravlenie* [Mechatronics, Automation, Control]. 2016, vol. 17, no. 1, pp. 3–10 (in Russ.).
5. *Analiz i optimalny sintez na EVM sistem upravleniya* [Analysis and Optimum Synthesis of Computer Control Systems]. A.A. Voronov, I.A. Orurk (Eds.). Moscow, Science, 1984, 344 p.
6. Goncharov V.I. *Veshchestvenny interpolyatsionny metod sinteza sistem avtomaticheskogo upravleniya* [Synthesis of electromechanical actuators for industrial robots]. Tomsk, TPU Publ., 2002, 100 p.
7. Than V.Z., Berchuk D.Yu. Experimental determination of the transfer function of the extruder heater. *Youth and modern IT: Proc. 13 Int. Science and Practice Conf. of Students, Postgraduates and Young Scientists*. Tomsk, 2015, pp. 261–262 (in Russ.).
8. Mikhalevich S.S., Baydali S.A., Chursin Yu.A. Graphical method for calculating regulatory settings that specifies the system to the parameters of stability. *Promyshlennyye ASU i kontrolyery* [Industrial ACS and controllers]. 2013, no. 10, pp. 35–39 (in Russ.).
9. Goncharov V.I., Liepinsh A.V., Rudnitsky V.A. The synthesis of low order robust controllers. *Izvestiya Ross. akademii nauk. Teoriya i sistemy upravleniya* [Jour. of Computer and Systems Sciences International]. 2001, no. 4, pp. 36–43.
10. Krutova I.N. Formation of a control algorithm for setting up iterative process in a system with a simplified reference model. *Avtomatika i telemekhanika* [Automation and Remote Control]. 1998, no. 2, pp. 72–84 (in Russ.).
11. Denisenko V.V. *Kompyuternoe upravlenie tekhnologicheskim protsessom, eksperimentom, oborudovaniem* [Computer control of a production process, experiment equipment]. Moscow, Goryachaya Liniya Telecom Publ., 2009, 608 p.
12. Goncharov V., Rudnicki V. Real interpolation method in automatic control systems self-adjustment problem. *Systems Science*. 2010, vol. 36, no. 3, pp. 35–37.

УДК 004.4:004.6::330.142.211
DOI: 10.15827/0236-235X.117.051-060

Дата подачи статьи: 01.09.16
2017. Т. 30. № 1. С. 51–60

ОБ ОДНОМ ПОДХОДЕ К РЕАЛИЗАЦИИ СИСТЕМЫ УПРАВЛЕНИЯ МАСТЕР-ДАННЫМИ ОБ АКТИВАХ

А.А. Сухобоков, к.т.н., artem.sukhobokov@yandex.ru

(Компания «САП СНГ», Космодамианская наб., 52/2, г. Москва, 115054, Россия);

В.И. Строгонова, инженер, vctrog@gmail.com

(Компания «Оптимальное Управление», ул. Фестивальная, 22, корп. 6, г. Москва, 125581, Россия)

В статье рассматриваются возможности современных MDM-систем (систем Master Data Management) и перспективные направления разработки мультидоменных и мультивекторных MDM-систем. Показаны причины, по которым однодоменные системы управления мастер-данными об активах не нашли широкого применения в отличие от существующих систем управления мастер-данными о клиентах, поставщиках, продуктах, сотрудниках и других типах бизнес-объектов. Рассматриваются сложности совмещения различных представлений одних и тех же активов в системе управления мастер-данными об активах. Делается вывод, что, пока не будут разработаны и не начнут успешно внедряться однодоменные системы управления мастер-данными об активах, переносить эту предметную область в мультидоменные системы преждевременно.

Для решения описанных проблем предложена модель мастер-данных об активах, позволяющая совместить их различные представления. Эта модель включает множество независимых иерархий для различных представлений одного и того же парка активов, неиерархические связи, специфические для каждой предметной области, решетки связей, позволяющие переходить между разными представлениями одного и того же актива, набор классификаторов активов, классы в которых определяют наборы описывающих активы атрибутов, классификаторы связей активов, а также структурные и функциональные модели отдельных типов активов. Для реализации модели мастер-данных об активах разработана архитектура MDM-системы, а также предложен алгоритм проверки корректности межракурсных связей всей модели в целом.

Выведены основные требования к инструментарию для разработки прототипа системы управления мастер-данными об активах: он должен одновременно обеспечивать функциональность графовой СУБД и графового энджина для выполнения сложных алгоритмов над графом в целом. Из двух существующих инструментов, отвечающих этим требованиям, для разработки выбран SAP HANA Graph.

Ключевые слова: MDM-система, система управления мастер-данными об активах, модель данных, архитектура системы, алгоритм проверки корректности связей, SAP HANA Graph.

Целями применения существующих систем управления мастер-данными (MDM-систем) являются формирование и поддержка консистентного, функционально полного и актуального представления основных данных, отражающих все нюансы структуры и деятельности предприятия. Наличие и поддержка такого единственного содержательного представления мастер-данных является критическим фактором для достижения бизнес-целей предприятия. MDM-системы представляют собой инструментарий для поддержки процессов управления мастер-данными в масштабе предприятия на основе принятых руководящих документов, политик и стандартов [1, 2]. С точки зрения принадлежности программных продуктов к прикладному или платформенному ПО MDM-системы входят в состав платформы и являются неотъемлемой частью комплекса систем управления информацией в масштабах предприятия (Enterprise Information Management) [3].

Системы управления мастер-данными об активах как подкласс систем управления мастер-данными

Ведущая мировая аналитическая компания Gartner традиционно ежегодно выпускает два независимых аналитических обзора по разным классам

MDM-систем. Эта традиция была сохранена и в 2015 году [4, 5]. Системы управления мастер-данными об активах отдельно не рассматриваются в этих обзорах. Они относятся к широкому классу MDM-систем с данными о предметах, и считается, что и на них распространяется обзор [4] (в отличие от MDM-систем с данными о командах, на которые распространяется обзор [5]).

Согласно [4], комплексная стратегия управления мастер-данными охватывает множество предметных областей: клиенты, продукты, материалы, услуги, активы, персонал или команды, поставщики и финансы. Мастер-данные об активах, как и мастер-данные о продуктах, становятся все более критичными в связи с ростом числа публикаций и интереса к Интернету вещей (IoT). Распространение IoT предполагает, что все больше устройств (вещей) обмениваются информацией. Многие из них находятся за пределами зоны действия используемых норм организации и управления информацией, в итоге теряется доверие к такой информации. Управление мастер-данными позволяет снять вопрос о доверии, обеспечить качество и консистентность данных об устройствах, а также данных, получаемых от этих устройств.

Рассматривая проблематику систем управления мастер-данными об активах, необходимо ответить на вопрос, что понимается под активом. Ответ не

всегда однозначен, и этот термин трактуется разными компаниями по-разному. Практически всегда в понятие активов включаются материальные активы (здания, сооружения, оборудование, земельные участки, строящиеся объекты и т.д.) и их составные части. Несколько реже в понятие включаются нематериальные активы – программные системы, БД, патенты, товарные знаки и т.д. Еще реже в российской практике [6] и более часто в западной в единое понятие активов включаются и финансовые активы – акции, облигации, займы, доли в совместных проектах, векселя и т.д.

Общая численность парка активов в крупных компаниях, таких как ОАО «РЖД», НК «Роснефть», может достигать нескольких десятков миллионов единиц. Основную часть такого большого парка (более 90 %), как правило, составляют материальные активы.

Обобщенно рассматривая системы управления мастер-данными об активах как некоторый подкласс MDM-систем о вещах, в [4] также уделяется много внимания перспективам работы MDM-систем с данными более чем одной предметной области (применительно к теме данной статьи – не только с мастер-данными об активах). Наряду с доминирующими сейчас однодоменными MDM-системами, предназначенными для работы с данными одной предметной области (например, с данными о клиентах или о продуктах), пока еще мало распространены, но быстро расширяются сегменты мультидоменных и мультивекторных MDM-систем. Мультидоменные системы предназначены для работы с данными нескольких предметных областей. Несмотря на то, что пользователи хотели бы иметь возможность работать с любыми предметными областями в любом месте и любым способом, существующие мультидоменные MDM-системы ориентированы на одну якорную область данных. В данный момент, согласно [4], мультидоменные MDM-системы – это:

- MDM-системы для работы с одной предметной областью, имеющие расширенные ссылки на другие предметные области;
- MDM-системы для работы с одной предметной областью, в которые добавлено большое число атрибутов и иерархий, чтобы поддержать локальные бизнес-процессы или приложения;
- несколько одновременно установленных MDM-систем для разных предметных областей, часто от разных поставщиков, объединенных внешним интерфейсом.

Концепция мультивекторных MDM-систем расширяет понятие мультидоменных MDM-систем. Помимо работы с разными предметными областями, мультивекторные MDM-системы должны работать в разных отраслях, при разных сценариях использования (проектирование структур, операционные или аналитические MDM-системы), для разных организационных структур (централизо-

ванная, федеративная, локализованная) и при разных сценариях внедрения MDM-системы (регистрационная, консолидационная, в режиме сосуществования и централизованная). По прогнозу Gartner [4], мультивекторные MDM-системы сменяют мультидоменные и однодоменные не ранее, чем через пять лет.

По мнению авторов, пока не будут реализованы и не докажут на реальных внедрениях свою работоспособность и эффективность однодоменные системы управления мастер-данными об активах, прогнозировать, что необходимые функциональные возможности по управлению мастер-данными об активах будут покрыты мультидоменными и мультивекторными MDM-системами, преждевременно. Только после того, как все необходимые механизмы будут отработаны на однодоменных системах, можно будет осуществить их перенос в мультидоменные системы.

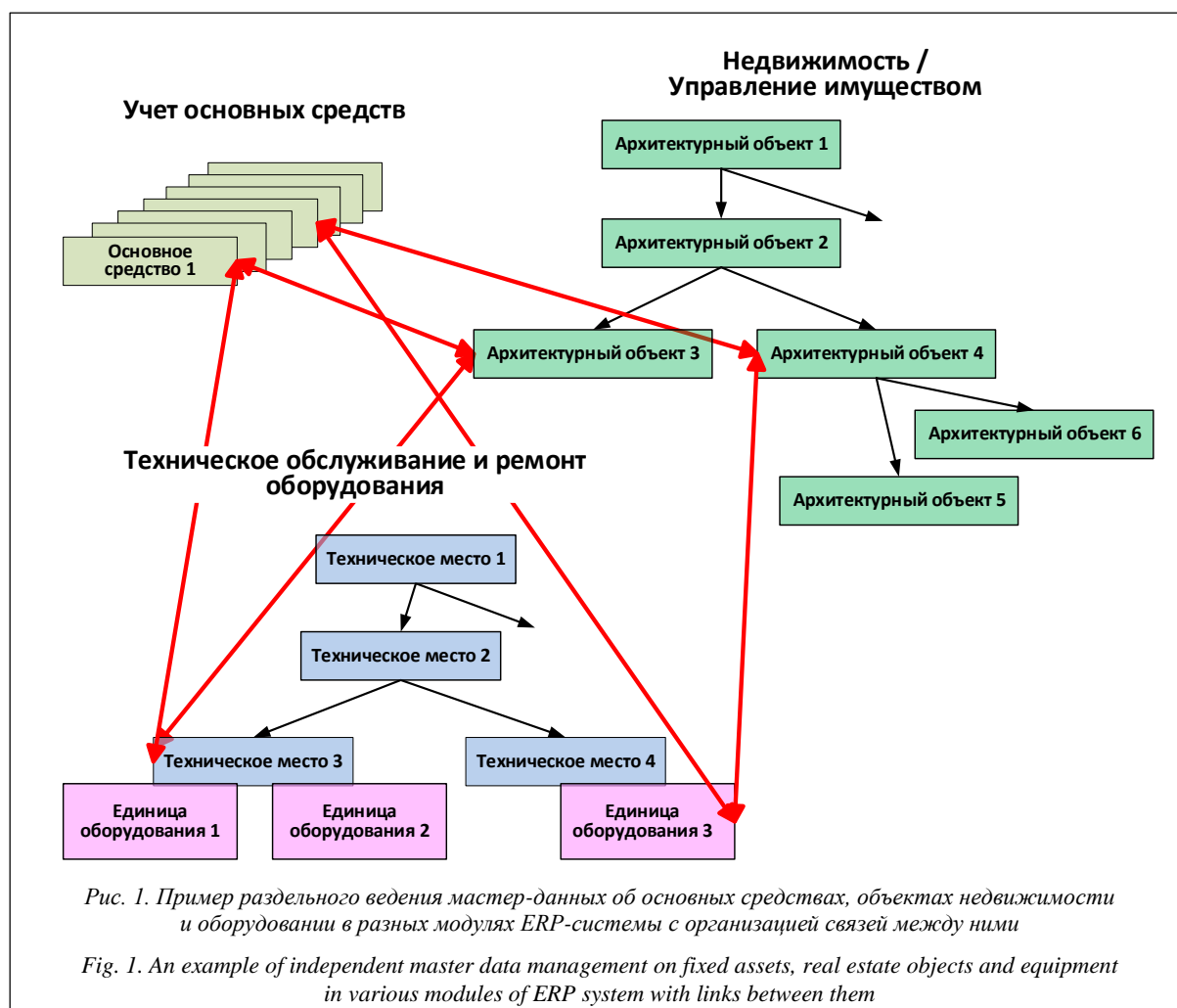
Нерешенные проблемы управления мастер-данными об активах

В существующих корпоративных системах управления в нескольких модулях или подсистемах независимо ведутся мастер-данные о производственных фондах (активах), отражающие различные характеристики одних и тех же объектов. Даже когда разные приложения являются модулями единой ERP-системы, используется такой же подход [7, 8]. Примерная структура основных данных об активах в существующих системах управления крупными компаниями показана на рисунке 1 [9].

Технически это реализуется следующим образом: в разных прикладных подсистемах для ведения мастер-данных об активах используются собственные структуры различной сложности и независимые классификаторы объектов. Мастер-данные ведутся разрозненно, хотя и перечень объектов, и значительная доля описывающих объекты характеристик (в том числе паспортных и эксплуатационных) необходимы для работы нескольких приложений. Как следствие, возникает дублирование данных и появляются противоречия между различными представлениями данных. Кроме снижения качества данных, это также увеличивает издержки по сопровождению [9].

В действительности в корпоративных системах имеется намного больше подсистем, в которых используются мастер-данные об активах, чем показано на рисунке 1 [9]. В их число входят данные:

- об основных средствах (объектах бухгалтерского и налогового учета);
- об объектах недвижимости;
- об объектах технического обслуживания и ремонта оборудования;
- об объектах имущества (правовой аспект);
- об объектах подсистем мониторинга (пожарной безопасности, видеонаблюдения и др.);



- об объектах подсистем оперативного управления режимами работы оборудования непрерывных производств, транспорта и связи, коммунального хозяйства;
- об объектах различных MES и технологических систем управления производством;
- об объектах геоинформационных систем (визуализируемых на картах);
- об объектах генерации, передачи и потребления электроэнергии в подсистемах управления производством и потреблением электроэнергии;
- об объектах добычи, транспортировки и отпуса нефти, нефтепродуктов и газа в подсистемах управления добычей, транспортировкой и сбытом углеводородов;
- об обрабатывающих центрах и транспортных механизмах в системах технологической подготовки производства (CAM-системах);
- о складах, распределительных центрах и торговых точках в системах управления розничными сетями.

Перечислим основные причины, по которым системы управления мастер-данными об активах не нашли широкого применения в отличие от MDM-систем для других предметных областей [9].

1. В разных прикладных областях применяются различные принципы выделения активов. Например, в рамках бухгалтерского учета в одно основное средство могут быть объединены несколько рядов стоящих единиц оборудования, имеющих одинаковые правила начисления амортизации. При проведении ремонта и технического обслуживания выделяются объекты, имеющие разный порядок ремонта и обслуживания. В рамках имущественного учета выделяются объекты, которые независимо могут быть проданы или сданы в аренду, например одно волокно в пучке оптических волокон или оптоволоконном кабеле.

2. В разных приложениях, работающих с одним и тем же парком активов, используются различные иерархии вложенности объектов, причем различаются объекты как группирующие уровни, так и нижних уровней таких детализирующих иерархий.

3. При функционировании комплексных систем управления пользователям очень часто необходимо видеть смежные представления одного и того же актива. Например, из технологического представления необходимо видеть объекты технического обслуживания и ремонта. Из объектов технического обслуживания и ремонта необходимо

видеть основные средства, на которые будут списываться затраты. Из основных средств необходимо видеть соответствующие объекты имущества и т.д. Почти всегда пользователям необходимо видеть объекты нескольких разных представлений. Отсутствие взаимно однозначного соответствия между объектами нижнего уровня различных приложений приводит к появлению связей многие ко многим между объектами разных представлений, описывающими один и тот же актив. Как уже было показано ранее, одному основному средству может соответствовать множество объектов технического учета. Но также справедливо и обратное: одному объекту технического учета, например асфальтированной дороге, может соответствовать множество основных средств, если ее отдельные участки, не являющиеся единицами технического учета, строились (выкупались) по разным ценам или в разное время.

Отсутствие в настоящий момент на рынке системы управления мастер-данными об активах, способной преодолеть перечисленные сложности, подтверждается также обзорами аналитической компании Forester. В первом квартале 2016 года компания выпустила обзор [10], в котором выделила четыре класса MDM-систем и проанализировала продукты двенадцати компаний. Были выделены

- MDM-системы, обеспечивающие интеграцию данных и решающие стандартные задачи;
- мультидоменные MDM-системы и MDM-системы, поддерживающие множество представлений мастер-данных;
- контекстуальные MDM-системы, поддерживающие семантическое представление бизнес-данных;
- аналитические MDM-системы, интегрированные внутри аналитических платформ.

Из четырех представленных классов MDM-систем сформулированные проблемы должны решаться в группе мультидоменных MDM-систем и MDM-систем, поддерживающих множество представлений мастер-данных. В качестве лидеров этой группы выделены SAS, SAP и IBM. У этих компаний в настоящий момент нет MDM-систем, способных обеспечить управление мастер-данными об активах в описанном выше ключе.

Предлагаемая модель данных

Для совмещения нескольких разных представлений одного и того же парка активов предлагается модель данных, включающая произвольное число ракурсов представления активов. Каждый ракурс описывается некоторой иерархией, примерный вид которой представлен на рисунке 2.

Начиная с некоторых узлов дальнейшие связи обрываются, чтобы не увеличивать сложность рисунка. Число уровней иерархии может быть раз-

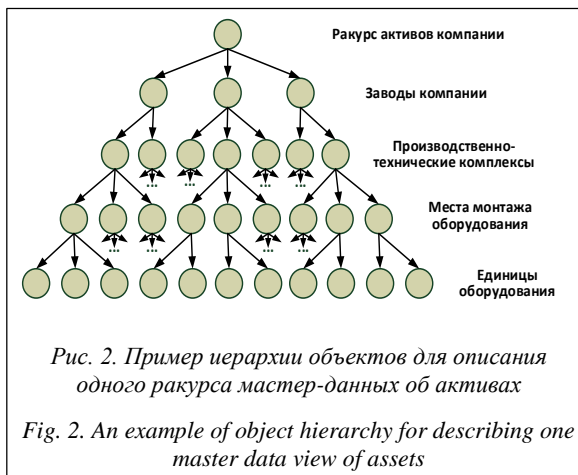


Рис. 2. Пример иерархии объектов для описания одного ракурса мастер-данных об активах

Fig. 2. An example of object hierarchy for describing one master data view of assets

личным в разных ракурсах. Иерархическая структура ракурса отображает принадлежность (вхождение объектов в объекты более высокого уровня).

Разные ракурсы полиморфны между собой, то есть отсутствует взаимно однозначное соответствие между объектами разных ракурсов.

Каждый объект, представленный в иерархии, принадлежит к какому-то классу объектов. Класс определяет свойства (набор атрибутов) объекта. Классы объектов строятся на принципах полиморфизма и наследования. Они образуют иерархию классов, позволяя детально описывать тип и атрибуты (характеристики) каждого актива. Для каждого ракурса строится собственная система классификации объектов. Атрибуты объектов могут иметь широкий спектр типов: не только целое, вещественное, логический признак, текст, но и массив, множество, зависимость, график, таблица, документ и др.

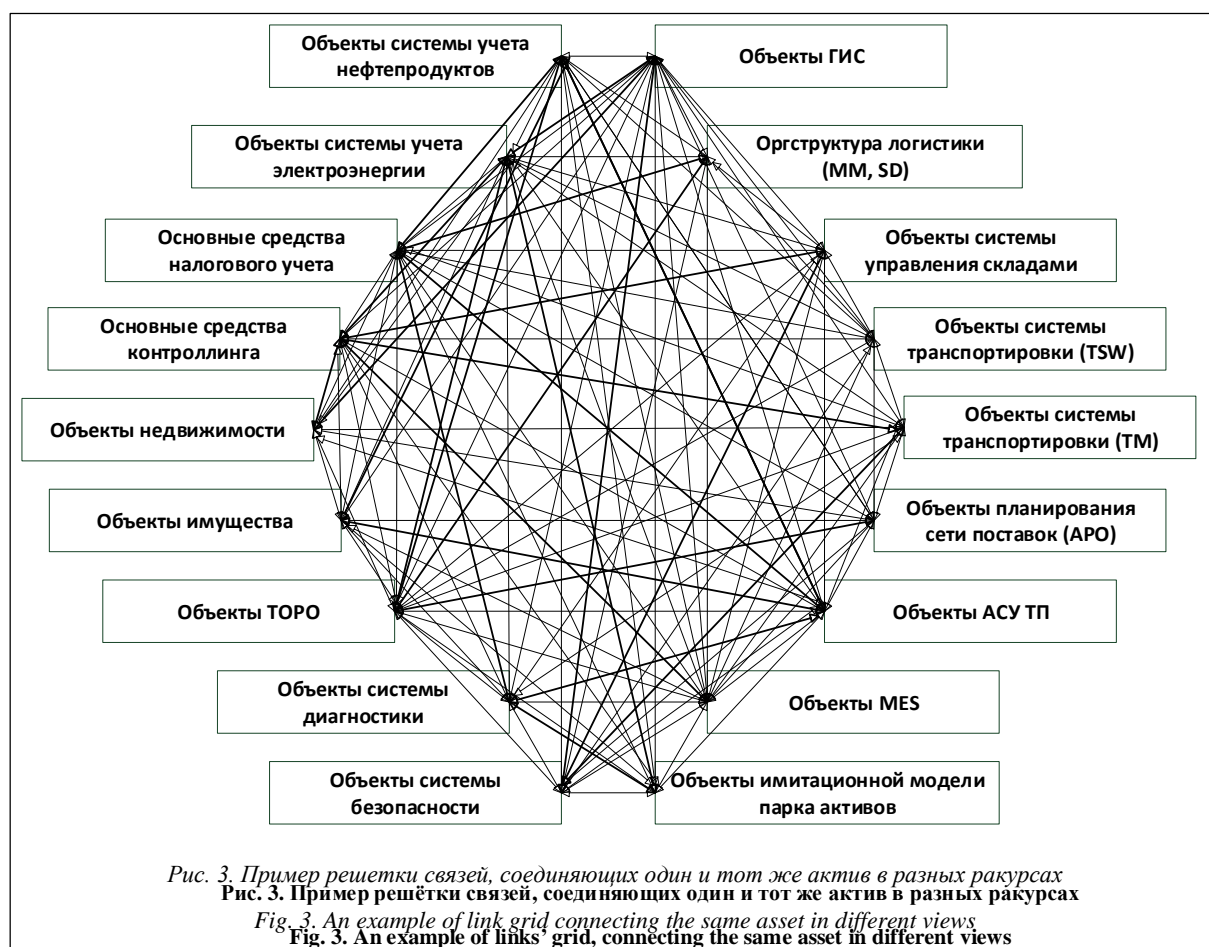
Для описания одного и того же актива, представленного в разных ракурсах, могут использоваться атрибуты с одинаковым именем, однако значения этих атрибутов могут не совпадать.

Параллельные иерархии разных ракурсов (лес) провязаны на каждом уровне решетками связей для обеспечения переходов между разными ракурсами представления одного и того же актива. Пример такой решетки представлен на рисунке 3.

Решетки необязательно являются полными (то есть необязательно в них присутствуют все связи каждого объекта с каждым). Наличие связей между разными объектами зависит от семантики: нужно устанавливать соотношения между объектами этих ракурсов на данном уровне или нет.

На одном уровне иерархии любого из ракурсов располагается много объектов, каждый из которых связан со своей решеткой смежных объектов.

Решетки одного уровня могут частично совмещаться друг с другом. Это возникает из-за того, что в отдельных ракурсах один объект может соответствовать нескольким объектам другого ракурса, расположенным на одном уровне иерархии (например, одно основное средство может объединять не-



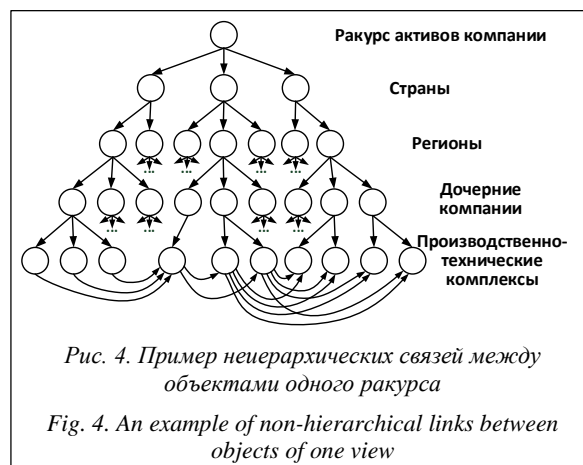
сколько единиц оборудования), и тогда решетки смежных объектов одного уровня могут частично соединяться в отдельных ракурсах.

Если рассматривать 3D-модель графа мастер-данных, то решетки на разных уровнях необязательно будут абсолютно параллельными. В отдельных ракурсах один объект может соответствовать нескольким объектам другого ракурса, расположенным на разных уровнях иерархии, и тогда решетки смежных уровней могут соединяться в отдельных ракурсах.

В отдельных ракурсах между объектами имеются неиерархические связи, отражающие специфические отношения между объектами, свойственные этому ракурсу. Например, функциональное взаимодействие в процессе производства, логистический обмен товарами, соседнее территориальное расположение и др. Пример ракурса с такими связями между производственно-техническими комплексами представлен на рисунке 4.

Показанные на рисунке 4 неиерархические связи могут иметь собственную систему классов и наборы атрибутов, описывающие каждую связь. В ряде случаев неиерархические связи между объектами одного ракурса могут быть параметрическими, то есть возникающими только тогда, когда значения одного или нескольких атрибутов объекта соответствуют определенным условиям.

Помимо классов объектов, определяющих их тип, в модели данных могут присутствовать структурные и функциональные модели отдельных типов активов. Одним из назначений таких моделей является проверка правильности вводимых мастер-данных. Возможности использования структурных описаний для проверки правильности ввода данных об активах можно проиллюстрировать следующим простым примером. Структурная модель котельной включает обязательные компоненты: здание, установленный в нем котел и дымовую трубу. Если вновь введенный пользователем объект имеет



тип «котельная», но не содержит всех трех обязательных компонентов, пользователю выводится сообщение об ошибке. Функциональные модели могут проверять соответствие объекта его типу, используя значения атрибутов этого объекта и атрибутов других, связанных с ним объектов. Структурные и функциональные модели, используемые для проверки полноты и правильности вводимых мастер-данных, могут быть достаточно сложными.

Роль структурных и функциональных моделей в общем механизме управления мастер-данными об активах не ограничивается проверкой правильности вводимых данных. При описании технических ракурсов представления активов в ряде случаев используемая схема соединения компонентов определяет тип объекта, состоящего из этих компонентов. Точно так же тип объекта может зависеть от значений его атрибутов и применяемых функциональных моделей.

Архитектура системы управления мастер-данными об активах

Предлагаемая общая архитектура системы управления мастер-данными об активах представлена на рисунке 5.

Ядро системы поддерживает создание/модификацию и хранение в памяти информационной модели активов.

Каждый интерфейсный модуль обеспечивает взаимодействие с конкретным приложением или функциональным модулем ERP-системы. По запросам, поступающим от внешних приложений или модулей ERP, интерфейсный модуль осуществляет выгрузку запрошенного фрагмента мастер-данных об активах, относящихся к конкретному ракурсу мастер-данных, с которыми работает запрашивающий модуль. Помимо объектов данных конкретного ракурса, могут быть выгружены также иерархия, классификатор, сетевые связи, структурные и функциональные модели, относящиеся к объектам этого ракурса. Все объекты активов могут быть выгружены вместе с имеющимися связями с объектами других ракурсов.

Помимо выгрузки из системы управления мастер-данными, интерфейсные модули также могут обеспечивать загрузку обновленных или новых мастер-данных соответствующих ракурсов из приложений и модулей ERP в систему управления мастер-данными. Однако такая загрузка не обеспечивает использование загруженных данных в системе. Требуется последующее подтверждение загруженных данных от администратора мастер-данных.

Модуль диалогового интерфейса обеспечивает создание и корректировку мастер-данных администратором мастер-данных в режиме диалога. В том числе с его помощью подтверждаются мастер-дан-

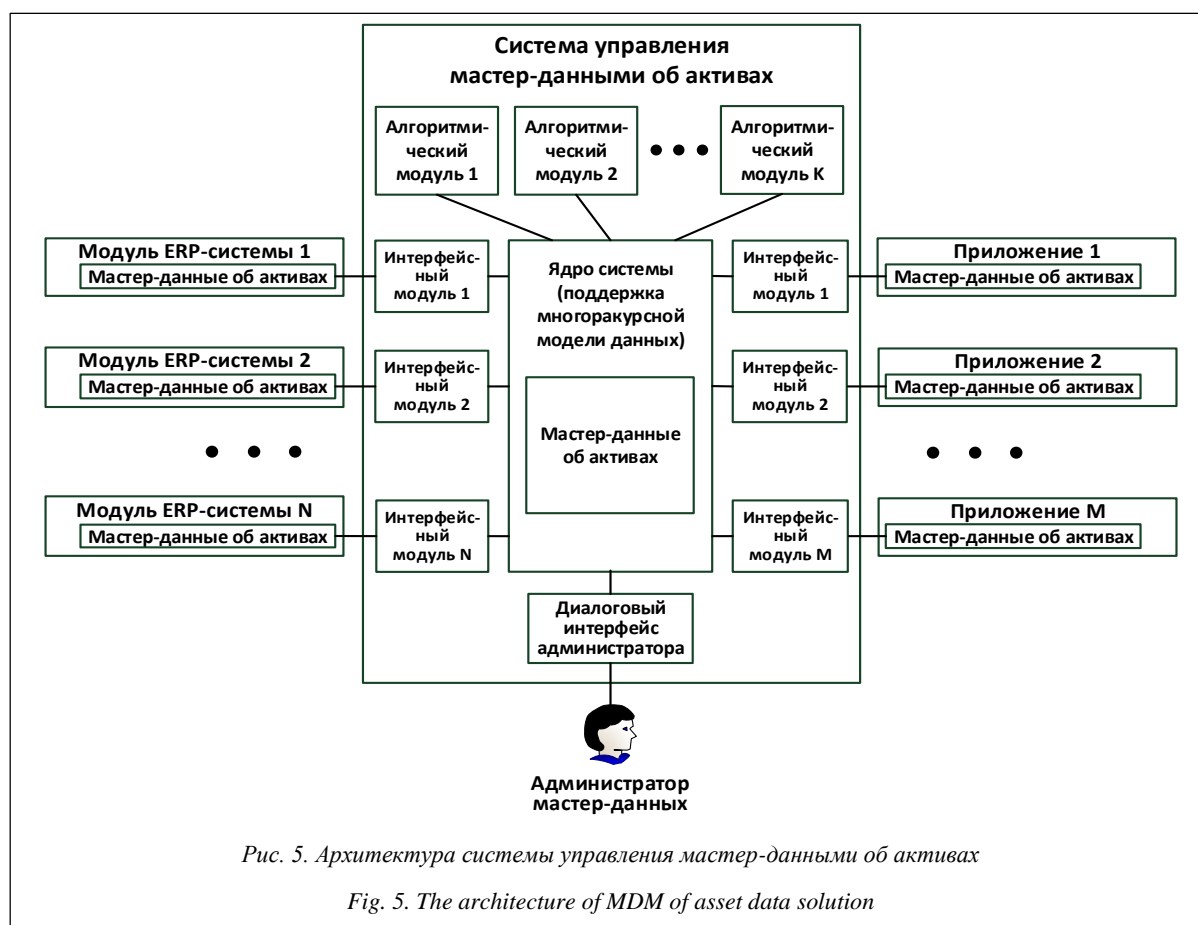


Рис. 5. Архитектура системы управления мастер-данными об активах

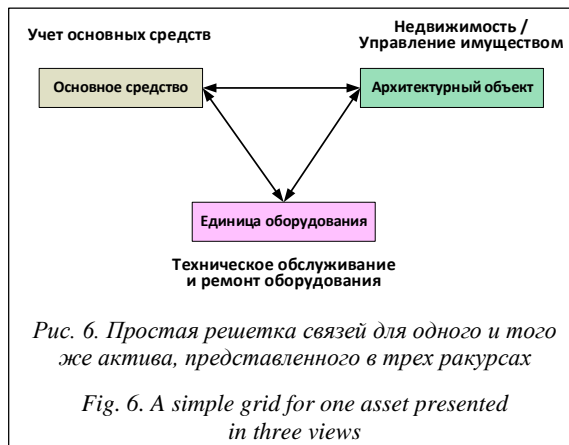
Fig. 5. The architecture of MDM of asset data solution

ные, загруженные интерфейсными модулями извне, или осуществляются корректировки этих данных, вносятся дополнительные связи и только потом дается разрешение на использование этих данных в системе.

Алгоритмические модули позволяют выполнить обработку всей модели мастер-данных в целом или отдельных ракурсов. С их помощью может выполняться проверка корректности имеющихся в системе мастер-данных или связей между объектами. Также с их помощью могут выполняться формирование новых ракурсов мастер-данных на основе имеющихся в системе данных других ракурсов и/или на основе внешних данных и другие операции над информационной моделью в целом. Инициирование работы алгоритмических модулей осуществляет администратор мастер-данных через модуль диалогового интерфейса.

Алгоритм проверки корректности межракурсных связей в единой модели активов

Для того чтобы единая модель активов была корректной, необходимо, чтобы в ней были корректными все связи. Для проверки структуры иерархий и неиерархических связей внутри ракурсов требуется использовать дополнительную семантическую информацию. В то же время проверку полноты и корректности всех решеток межракурсных связей можно выполнить, используя только тип объектов и существующую структуру связей без привлечения дополнительной информации о конкретных объектах. Для проверки полноты связей в решетке можно использовать шаблон, содержащий все необходимые связи, и каждую решетку можно сверить с этим шаблоном. Поскольку в больших компаниях неизбежно будет наблюдаться значительное разнообразие видов активов, большинство из которых не должно присутство-



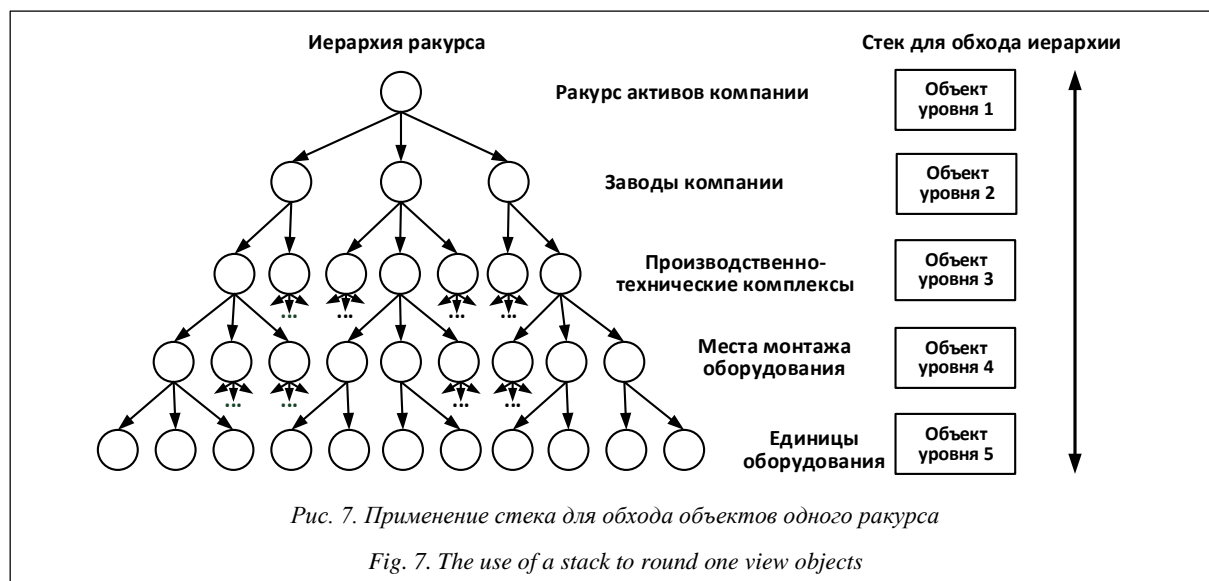
вать во всех ракурсах, шаблон решетки не может быть единым, на практике должен использоваться некоторый набор правил, регламентирующих применение шаблонов для отдельных видов активов.

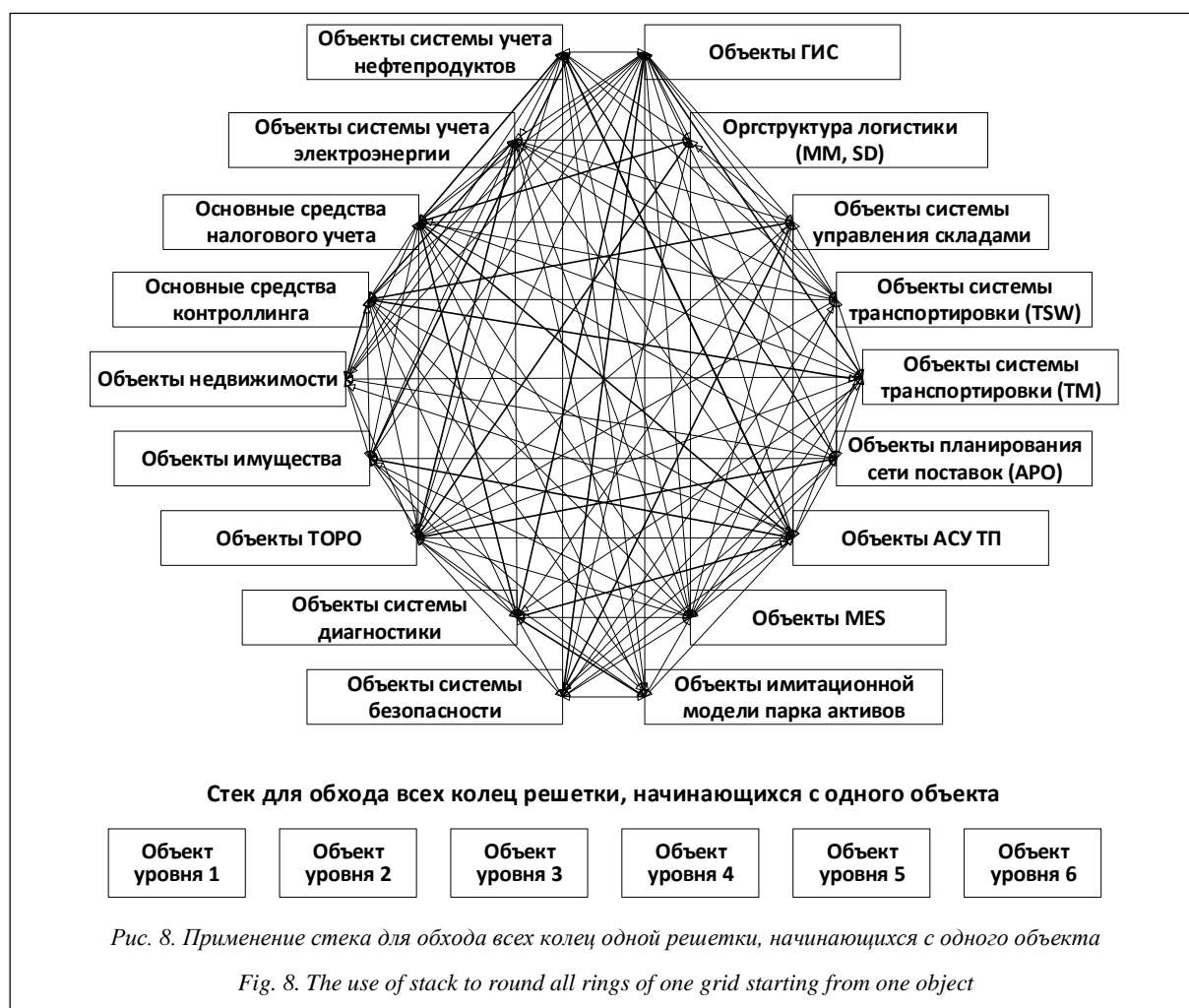
Однако наличие конечного множества используемых шаблонов решеток и сверка с ними фактических решеток связей не избавит от ситуации, когда связи между разными объектами требуемых типов перепутаны между собой. Для борьбы с этим явлением авторы предлагают специальный алгоритм, рассматриваемый далее.

Если предположить, что имеются всего три ракурса активов, решетка принимает вид, представленный на рисунке 6.

Проверка корректности всех связей решетки в этом случае осуществляется с помощью простого обхода: из каждого объекта необходимо пройти по кольцу связей в одном направлении, а потом в обратном. Если для всех объектов при проходе в обоих направлениях мы возвращаемся к первоначальному объекту, все связи в решетке корректны.

Обобщим этот подход для сложных многоракурсных структур описания активов. Для обхода всех объектов в иерархии каждого ракурса используем стек, показанный на рисунке 7.





В процессе проверки необходимо обойти все иерархические ракурсы, представленные в структуре активов, и запустить алгоритм проверки связей для каждого объекта.

Для обхода всех колец одной решетки, начинающихся с одного объекта, используем еще один стек, представленный на рисунке 8.

Пройдя по каждому из колец, необходимо вернуться к первоначальному объекту. Поскольку осуществляется перебор всех связей каждого объекта, автоматически будет реализован просмотр в прямом и обратном направлениях. Дополнительно программировать проход в обратном направлении не требуется.

Чтобы не проверять многократно одни и те же варианты колец связей, начиная с разных объектов, на время выполнения алгоритма вводится разметка связей. Если при обходе колец проходим по какой-то связи и для следующего перехода видим только одну связь, у которой нет отметки, или конец обхода кольца, то ставим на пройденной связи отметку. При дальнейшей проверке проходить по этой связи не нужно. В результате после обхода всех колец, начиная с объектов одного ракурса, если объекты другого ракурса не имеют допол-

нительных связей, которые еще не проверялись, обход этих объектов не потребует какого-либо обхода колец. Обобщенная блок-схема описанного алгоритма представлена на рисунке 9.

Использование полученных результатов при разработке прототипа системы управления мастер-данными об активах

Исходя из особенностей разработанной модели данных и алгоритма проверки корректности меж-ракурсных связей, при разработке прототипа системы управления мастер-данными об активах сформулированы следующие основные требования к используемому инструментарию.

1. Инструментарий должен обладать возможностями работы с графовой БД [11]. Это обеспечит существенные преимущества при работе с графами за счет увеличения скорости разработки и сокращения объема кода при обходе цепочек связей по сравнению с традиционным SQL, а также будет требоваться меньше ресурсов для работы прототипа.

2. Инструментарий должен обладать возможностями графовых энджинов [12], выполняющих

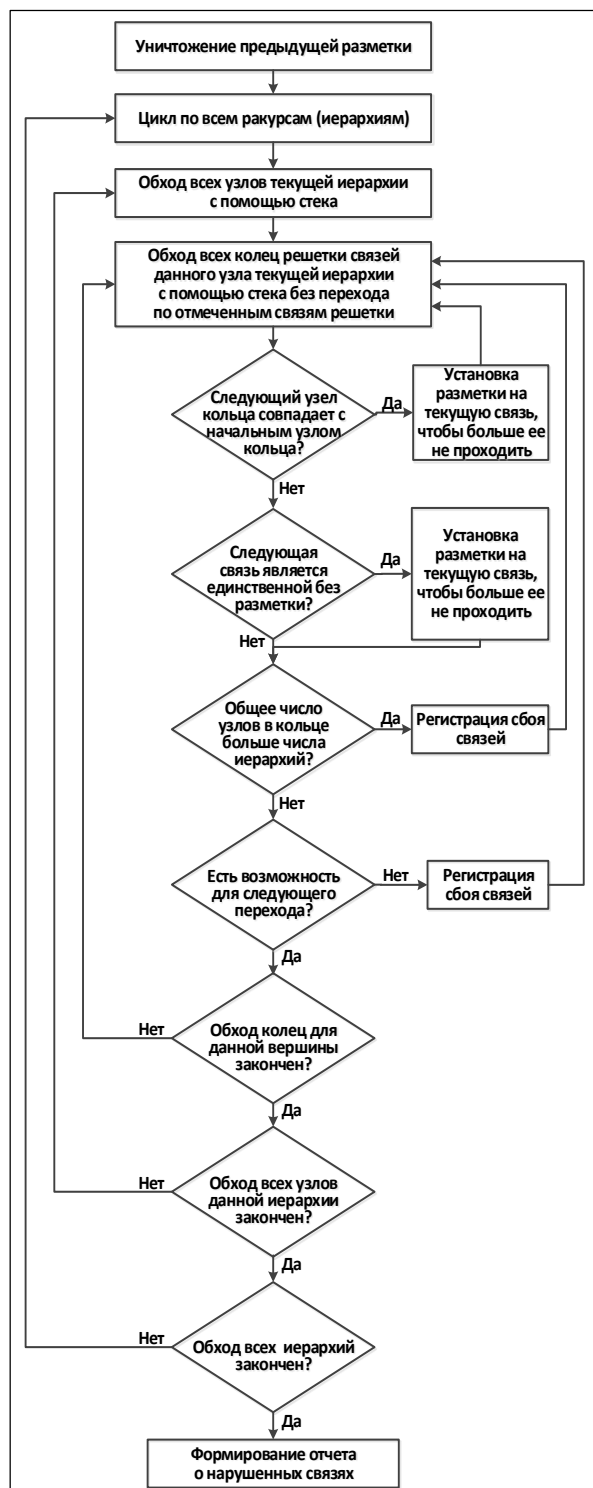


Рис. 9. Обобщенная блок-схема алгоритма проверки целостности межракурсных связей в единой модели активов

Fig. 9. The generalized block diagram of the algorithm for verification links integrity between views of the single asset model

сложные алгоритмы над большим графом в целом, даже если его фрагменты хранятся на разных серверах кластера.

Этим требованиям соответствуют SAP HANA Graph [13] и Microsoft Graph Engine [14, 15], работающий на кластере серверов под управлением Microsoft Windows Server. Для использования Microsoft Graph Engine требуется специально конфигурировать кластер под решаемую задачу, в то время как SAP HANA Graph может работать на любой SAP HANA, установленной в режиме On-Premise. Поэтому для использования выбран SAP HANA Graph. Перед выполняемой в настоящее время разработкой прототипа системы управления мастер-данными об активах ставятся цели обеспечить поддержку разработанной модели данных, реализовать механизм проверки полноты межракурсных связей и разработанный алгоритм проверки корректности межракурсных связей.

Литература

1. Wolter R., Haselden K. The what, why, and how of master data management. 2006. URL: <https://msdn.microsoft.com/en-us/library/bb190163.aspx?f=255&MSPPError=-2147217396> (дата обращения: 29.07.2016).
2. Otto B., Schmidt A. Enterprise master data architecture: design decisions and options. URL: http://mitiq.mit.edu/ICIQ/Documents/IQ%20Conference%202010/Papers/2B1_EnterpriseMasterDataArchitecture.pdf (дата обращения: 29.07.2016).
3. Barrenechea M.J., Jenkins T. Enterprise information management: the next generation of enterprise software. Open Text Corporation, Waterloo (Canada), 2013, 288 p.
4. Magic quadrant for master data management of product data solutions. Gartner, November 12, 2015, ID:G00271935.
5. Magic quadrant for master data management of customer data solutions. Gartner, November 11, 2015, ID:G00271783.
6. Система управления // Годовой отчет ПАО «ЕЭС России». 2006. С. 26–30. URL: http://www.fsk-ees.ru/media/File/stockholders/otchet/decisions/VOSA_otchet/RAO/03_RAO_GO_2006.pdf (дата обращения: 29.07.2016).
7. Equipment as units of tangible assets. URL: http://help.sap.com/saphelp_ppm400/helpdata/en/01/d545f84ab311d189740000e8322d00/frameset.htm (дата обращения: 29.07.2016).
8. Creating an architectural object. URL: http://help.sap.com/saphelp_erp60_sp/helpdata/en/fb/5ad0531d8b4208e10000000a174cb4/frameset.htm (дата обращения: 29.07.2016).
9. Система управления мастер-данными об активах. URL: <http://optimalmngmnt.com/page34ru.htm> (дата обращения: 29.07.2016).
10. The Forrester Wave™: Master Data Management, Q1 2016. Forester, March 16, 2016, 17 p.
11. Robinson I., Webber J., Eifrem E. Graph databases. O'Reilly Media, 2nd ed., 2015, 238 p.
12. Сакр Ш. Обработка больших объемов графовых данных: путеводитель по современным технологиям. URL: <http://www.ibm.com/developerworks/ru/library/os-giraph/index.html> (дата обращения: 29.07.2016).
13. SAP HANA graph reference. URL: http://help.sap.com/hana/SAP_HANA_Graph_Reference_en.pdf (дата обращения: 29.07.2016).
14. Shao B., Liu T.-Y., Ma W.-Y., Li Y. Graph engine, May 14, 2015. URL: <https://www.microsoft.com/en-us/research/project/graph-engine/> (дата обращения: 29.07.2016).
15. Graph engine. URL: <http://www.graphengine.io/#summary> (дата обращения: 29.07.2016).

ON AN APPROACH TO CONSTRUCT ASSET MASTER DATA MANAGEMENT SYSTEM

A.A. Sukhobokov¹, Ph.D (Engineering), artem.sukhobokov@yandex.ru
V.I. Stroganova², engineer, vctrog@gmail.com

¹ SAP CIS, Kosmodamianskaya Emb. 52/2, Moscow, 115054, Russian Federation

² Optimal Management, LLC, Festivalnaya St. 22/6, Moscow, 125581, Russian Federation

Abstract. The paper describes capabilities of current MDM (Master Data Management) solutions and prospects of multidomain and multivector MDM solutions. The paper presents the reasons for which MDM of asset data solutions for single data domain are not used successfully in contrast to the existing MDM solutions for other data domains such as customers, suppliers, products, employees, etc. There are challenges of combining several different representations of the same assets in MDM of asset data solution. The conclusion shows that as long as relevant single-domain MDM of asset data solutions are not developed and not implemented successfully, it is too early to move this subject area to multidomain systems.

To solve the problems described above, the authors propose a model of assets master data, which enable to combine different representations. This model includes multiple independent hierarchies for different representations of the same assets, non-hierarchical links specific for each subject area, grids of links allowing to go between different representations of the same asset, a set of asset classifiers whose classes define sets of attributes for describing assets, classifiers of links between assets, as well as structural and functional models for individual asset types. In order to implement the proposed model of master data on assets, the authors have developed a special architecture of MDM of asset data solution, as well as an algorithm for checking the integrity of links between different representations across the whole data model.

Key requirements are defined to the tools for developing a prototype of MDM of asset data solution. It must provide the functionality of a graph DBMS and at the same time a graph engine to perform complex algorithm on the graph as a whole.

Keywords: MDM solution, master data management of asset data solution, data model, system architecture, links integrity checking algorithm, SAP HANA Graph.

References

1. Wolter R., Haselden K. *The What, Why, and How of Master Data Management*. Microsoft Corporation, November 2006. Available at: <https://msdn.microsoft.com/en-us/library/bb190163.aspx?f=255&MSPPErr=-2147217396> (accessed July 29, 2016).
2. Otto B., Schmidt A. *Enterprise Master Data Architecture: Design Decisions and Options*. Institute of Information Management, University of St. Gallen, 13 p. Available at: http://mitiq.mit.edu/ICIQ/Documents/IQ%20Conference%202010/Papers/2B1_EnterpriseMasterDataArchitecture.pdf (accessed July 29, 2016).
3. Barrenechea M.J., Jenkins T. *Enterprise Information Management: The Next Generation of Enterprise Software*. Open Text Corporation, Waterloo Canada, 2013, 288 p.
4. *Magic Quadrant for Master Data Management of Product Data Solutions*. Gartner. November 12, 2015, ID:G00271935.
5. *Magic Quadrant for Master Data Management of Customer Data Solutions*. Gartner. November 11, 2015, ID:G00271783.
6. *Management system. Annual report RAO UES*, 2006, pp. 26–30. Available at: http://www.fsk-ees.ru/media/File/stockholders/otchet/decisions/VOSA_otchet/RAO/03_RAO_GO_2006.pdf (accessed July 29, 2016) (in Russ.).
7. Equipment as Units of Tangible Assets. *SAP Help Portal*. Available at: http://help.sap.com/saphelp_ppm400/helpdata/en/01/d545f84ab311d189740000e8322d00/frameset.htm (accessed July 29, 2016).
8. Creating an Architectural Object. *SAP Help Portal*. Available at: http://help.sap.com/saphelp_erp60_sp/helpdata/en/fb/5ad0531d8b4208e10000000a174cb4/frameset.htm (accessed July 29, 2016).
9. *System of Asset Master Data Management*. Optimal Management LLC, Available at: <http://optimalmngmnt.com/page34en.htm> (accessed July 29, 2016).
10. *The Forrester Wave™: Master Data Management, Q1 2016*. Forester, March 16, 2016, 17 p.
11. Robinson I., Webber J., Eifrem E. *Graph Databases*. 2nd ed., O'Reilly Media Publ., 2015, 238 p.
12. Sakr S. *Processing large-scale graph data: A guide to current technology*. IBM. Available at: <http://www.ibm.com/developerworks/library/os-giraph/> (accessed July 29, 2016).
13. *SAP HANA Graph Reference*. SAP HANA Platform SPS 12, Document Version: 1.0 – 2016-05-11, 28 p. Available at: http://help.sap.com/hana/SAP_HANA_Graph_Reference_en.pdf (accessed July 29, 2016).
14. Shao B., Liu T.-Y., Ma W.-Y., Li Y. *Graph Engine*. Microsoft Research, 2015. Available at: <https://www.microsoft.com/en-us/research/project/graph-engine/> (accessed July 29, 2016).
15. *Graph Engine*. Microsoft. Available at: <http://www.graphengine.io/#summary> (accessed July 29, 2016).

УДК 519.767.2

DOI: 10.15827/0236-235X.117.061-066

Дата подачи статьи: 13.12.16

2017. Т. 30. № 1. С. 61–66

ПОСТРОЕНИЕ МОДЕЛЕЙ СИСТЕМ НА БАЗЕ ЭКВАЦИОНАЛЬНОЙ ХАРАКТЕРИСТИКИ ФОРМУЛ LTL

Ю.П. Кораблин, д.т.н., профессор, y.p.k@mail.ru
(Российский государственный социальный университет,
ул. Вильгельма Пика, 4, г. Москва, 129256, Россия);
А.А. Шипов, ведущий инженер, a-j-a-1@yandex.ru
(Московский технологический университет (МИРЭА),
просп. Вернадского, 78, г. Москва, 119454, Россия)

Задача верификации как программных, так и технических систем всегда была и остается одной из самых значимых с момента появления первых вычислительных устройств. Сегодня уже существует достаточно большое количество подходов к решению данной проблемы. Однако именно развитие такого метода формальной верификации, как Model Checking, позволило решить проблему представления верифицируемых систем и унифицировать процесс верификации программных и технических систем. Идея, лежащая в его основе, состоит в приведении исходной системы к некоторой унифицированной форме, то есть для выполнения верификации требуется лишь наличие модели, которая максимально точно описывала бы ее поведение.

В данной статье рассматривается возможность построения моделей систем с помощью RLTL-нотации (Recursive Linear Temporal Logic), которая является рекурсивным представлением формул логики линейного времени. Однако ее использование не ограничивается только этим аспектом.

Преимущество от использования для этих целей именно RLTL состоит в том, что модели, заданные с ее помощью, могут быть верифицированы относительно требований, заданных на основе RLTL, без перехода к другим структурам данных, что, безусловно, будет способствовать упрощению и повышению быстродействия процесса верификации. Кроме того, в работе предложены формальные средства, позволяющие во многих случаях существенно упростить модель, построенную на основе RLTL, за счет сокращения числа ее состояний и переходов.

Ключевые слова: верификация, Model Checking, эквивалентная характеристика RLTL, модель Крипке, автомат Бюхи, формула временной логики, LTL, CTL.

Верификация работы программных и технических систем является неотъемлемой частью их жизненного цикла. Существует целый ряд инструментов и средств верификации, которые являются более удобными и эффективными при тех или иных условиях, на тех или иных этапах проектирования и сопровождения. Однако за последние два десятилетия лучше всего себя зарекомендовали инструменты формальной верификации, работа которых может быть автоматизирована, а корректность функционирования – доказана формально. Наиболее активно используемым инструментом на сегодняшний день является метод формальной верификации на моделях, или Model Checking [1, 2].

Технология процесса верификации программных и технических систем методом проверки на моделях подразумевает создание моделей систем, свойства которых требуется проверить. Так, например, модель системы может быть задана с помощью различных структур данных, в частности таких, как структура Крипке, язык асинхронных процессов Promela, в виде конечного автомата и т.д. Однако независимо от представления модели системы и верифицируемых свойств, заданных на базе логики линейного времени (LTL) [3, 4], необходимо, как правило, процесс их преобразования к автоматам Бюхи. При этом алгоритмы преобразования формул LTL в автоматы Бюхи являются нетривиальными, а принципы их функционирования непонятны на интуитивном уровне.

В данной статье рассматривается новая нотация RLTL (recursive linear temporal logic) для представления формул LTL в виде системы рекурсивных уравнений с целью расширения ее описательной мощности [5]. Кроме того, RLTL может быть использована и для задания моделей верифицируемых систем, что предоставляет следующие важные преимущества:

- задание как модели самой системы, так и проверяемых относительно нее свойств на базе единой структуры данных;
- возможность быстрого и интуитивно понятного преобразования RLTL-структур в автоматы Бюхи для их последующей верификации;
- возможность выполнения верификации без преобразования к автомату Бюхи;
- ускорение процесса верификации, что крайне важно при верификации больших распределенных программных систем.

RLTL-нотация

Рассмотрим рекурсивные представления основных базовых операторов LTL (табл. 1).

Символом \circ обозначен оператор продолжения (конкатенация выражений), что позволяет использовать оператор X в неявном виде, а также упростить его восприятие, в частности, запись $\varphi_1 \circ \varphi_2$ будет пониматься как « φ_2 следует за φ_1 ». Под символом Δ , далее именуемым предикатом неопреде-

ленности, будем понимать некоторое неопределенное подмножество символов входного алфавита. Под отрицанием предиката неопределенности будем понимать некоторое другое неопределенное подмножество символов алфавита, а под отрицанием отрицания – некоторое третье неопределенное подмножество. Таким образом, отрицание предиката неопределенности является неполным, а каждое последующее отрицание дает неопределенное подмножество символов. Каждое из этих подмножеств ввиду своей неопределенности также может быть обозначено через Δ .

Таблица 1

Рекурсивные представления операторов LTL

Table 1

Recursive presentations of LTL operators

LTL	RLTL
$F\phi = \phi \vee XF\phi$	$F\phi = \phi \vee \Delta \circ F\phi$
$G\phi = \phi \wedge XG\phi$	$G\phi = \phi \circ G\phi$
$\phi_1 U \phi_2 = \phi_2 \vee \phi_1 \wedge X(\phi_1 U \phi_2)$	$U(\phi_1, \phi_2) = \phi_2 \vee \phi_1 \circ U(\phi_1, \phi_2)$
$\phi_1 \wedge X\phi_2$	$\phi_1 \circ \phi_2$
$X\phi$	$\Delta \circ \phi$

В таблице 1 представлены почти все операторы, используемые в LTL, кроме оператора $X\phi$, у которого отсутствует рекурсивная форма. Таким образом, поскольку все основные операторы LTL могут быть представлены в рекурсивной форме, авторы полагают, что приведение их к некоторому унифицированному виду, который, следуя [5], будем именовать эквациональной характеристикой, позволит существенно расширить описательную способность LTL и упростит процесс верификации ее формул.

Сочетания операторов LTL также могут быть выведены с помощью аксиом в рамках эквациональной характеристики. В таблице 2 приведены некоторые сочетания операторов LTL.

Таблица 2

Сочетания операторов LTL в эквациональной форме

Table 2

Equational combinations of LTL operators

LTL	RLTL
$FG\phi$	$F' = \phi \circ G' \vee \Delta \circ F'$ $G' = \phi \circ G'$
$GF\phi$	$G' = \phi \circ G' \vee \Delta \circ F'$ $F' = \phi \circ G' \vee \Delta \circ F'$

Сформулируем основные аксиомы логики RLTL:

A1.	$FF\phi = F\phi$	(FF)
A2.	$GG\phi = G\phi$	(GG)
A3.	$\neg F\phi = G\neg\phi$	(\neg F)
A4.	$\neg G\phi = F\neg\phi$	(\neg G)
A5.	$G\phi = \phi \circ G\phi$	(G)
A6.	$F\phi = \phi \vee \Delta \circ F\phi$	(F)
A7.	$G\phi \vee \neg G\phi = \Delta$	(\vee G)
A8.	$F\phi \vee \neg F\phi = \Delta$	(\vee F)
A9.	$G\phi_1 \circ \phi_2 = G\phi_1$	(G \circ)
A10.	$F\phi_1 \circ \phi_2 = F(\phi_1 \circ \phi_2)$	(F \circ)

A11.	$U(\phi_1, \phi_2) = \phi_2 \vee \phi_1 \circ U(\phi_1, \phi_2)$	(U)
A12.	$\neg(\phi \circ \phi_1) = \neg\phi \vee \Delta \circ \neg\phi_1$	(\neg \phi \circ)
A13.	$\neg\Delta \rightarrow \Delta$	(\neg\Delta)
A14.	$\Delta \wedge \phi = \phi$	(\Delta \wedge)
A15.	$\phi = \phi \circ \Delta$	(\phi \Delta)

Построение RLTL-моделей

Структуры на базе RLTL-нотации могут быть интерпретированы в качестве недетерминированного конечного автомата, который на вход получает бесконечные цепочки и моделирует поведение реагирующих систем. Так, рассмотрим формальное определение автомата Бюхи. Автомат Бюхи – это пятерка $(S, \Sigma, S_0, \delta, F)$, где Σ – конечное множество символов (алфавит); S – множество состояний; $S_0 \subseteq S$ – множество начальных состояний (содержащее только один элемент для детерминированного случая); $\delta: S \times \Sigma \rightarrow 2^S$ – отношение переходов (для детерминированного автомата $\delta: S \times \Sigma \rightarrow S$); $F \subseteq S$ – множество допускающих (финальных) состояний.

Нетрудно убедиться, что любая структура на базе RLTL-нотации может быть интерпретирована в качестве автомата Бюхи в соответствии с данным определением. В качестве алфавита в RLTL также выступает некоторое множество символов Σ . В качестве состояний выступает множество метаварiableных $S = \{F_1, F_2, \dots, F_n\}$, задающих уравнения системы. По умолчанию в качестве начального состояния выступает метаварiableная первого уравнения системы, однако, если существует несколько начальных состояний, их можно задать в виде элементов соответствующего множества. Функция переходов для RLTL остается аналогичной автоматам Бюхи, где в роли переходов выступает оператор продолжения \circ , за которым следует метаварiableная, обозначающая состояние, в которое осуществляется переход. Что касается множества допустимых или финальных состояний, то в RLTL не существует аналога, поэтому необходимо явное задание множества соответствующих метаварiableных.

Для задания с помощью RLTL-нотации моделей систем достаточно использовать лишь тройку значений (S, Σ, δ) , поскольку каждое состояние модели системы является допустимым, а любое множество начальных состояний из более чем одного элемента может быть заменено единственным начальным состоянием, из которого существуют переходы по любому из символов алфавита в каждое из этих состояний. В качестве примера зададим в терминах RLTL простую модель работы светофора на основе ее устного описания: «Горит зеленый сигнал, затем загорается желтый, а после желтого – красный. После красного сигнала снова загорается желтый, за ним зеленый. В такой последовательности сигналы продолжают чередоваться дальше». Опишем модель работы светофора следующей тройкой значений (S, Σ, δ) : $S = \{F_1, F_2, F_3, F_4\}$, $\Sigma = \{g, y, r\}$, $\delta = \{g: F_1 \rightarrow F_2, y: F_2 \rightarrow F_3, r: F_3 \rightarrow F_4, y: F_4 \rightarrow F_1\}$, где g – зеленый сигнал светофора; y – жел-

тый сигнал; r – красный сигнал. В RLTL-нотации модель будет иметь следующий вид: $F_1 = g \circ F_2$, $F_2 = y \circ F_3$, $F_3 = r \circ F_4$, $F_4 = y \circ F_1$.

Использование пятерки значений ($S, \Sigma, S_0, \delta, F$) требуется лишь в тех случаях, когда необходимо превратить данные на базе RLTL в автомат Бюхи. Так, пусть дана формула LTL $GF\varphi$ – «всегда в будущем будет наступать φ ». В терминах RLTL-нотации она будет выглядеть следующим образом: $F_1 = \varphi \circ F_1 \vee \Delta \circ F_2$, $F_2 = \varphi \circ F_1 \vee \Delta \circ F_2$.

В данном случае система задана тройкой значений (S, Σ, δ), где $S = \{F_1, F_2\}$, $\Sigma = \{\varphi\}$, $\delta = \{\varphi: F_1 \rightarrow F_1, \varphi: F_2 \rightarrow F_1, \Delta: F_1 \rightarrow F_2, \Delta: F_2 \rightarrow F_2\}$. Для превращения данной формулы в автомат Бюхи достаточно указать множество начальных и множество допустимых состояний: $S_0 = \{F_1\}$ и $F = \{F_1\}$. Аналогично может быть задана и модель любой системы на базе RLTL.

Таким образом, любая RLTL-структура, использующая тройку значений (S, Σ, δ), может быть превращена в автомат Бюхи за счет задания множеств начальных и допустимых состояний. Возможность простого и интуитивно понятного превращения формул RLTL в автоматы Бюхи и наоборот, как и представленные выше примеры, подтверждает возможность задания моделей верифицируемых систем сразу на базе RLTL. Сам процесс задания моделей систем может быть выполнен двумя способами:

- построение модели системы на базе структуры Крипке с ее последующим преобразованием в RLTL; поскольку структуры Крипке могут быть быстро преобразованы в автоматы Бюхи [1], их преобразование в RLTL-формулы также выполняется относительно просто;

- непосредственное построение моделей систем сразу на базе RLTL-нотации, что является более приемлемым, поскольку дальнейшие преобразования для их верификации не требуются.

Абстракция и унификация RLTL-моделей

Для повышения быстродействия процесса верификации больших распределенных систем существует ряд специальных инструментов, средств и алгоритмов:

- редукция частичных порядков [6];
- абстракция данных [6];
- хэширование битовых состояний (вместо полных состояний хранится лишь их хэш, что снижает требования к объему памяти, но уменьшает полноту системы) [7];
- ускорение проверки «корректности в слабом смысле», или Weak Fairness [8].

Использование данных средств позволяет снизить проблему комбинаторного взрыва, то есть быстрого роста числа состояний.

Для упрощения моделей систем введем операцию абстракции и унификации. Операция абстрак-

ции позволяет осуществлять сокращение числа свойств моделей систем, а операция унификации – объединение однотипных элементов. Операцию абстракции обозначим как ABS_H^α , где H – множество заменяемых символов; α – заменяющий символ (тот символ, на который заменяются элементы из множества H); φ – атомарный предикат (символ алфавита); μ_i – формула. Аксиомы, описывающие свойства данной операции, представлены далее:

$$AB1 \quad ABS_H^\alpha(\varphi) = \varphi, \text{ если } \varphi \notin H$$

$$AB2 \quad ABS_H^\alpha(\varphi) = \alpha, \text{ если } \varphi \in H$$

$$AB3 \quad ABS_H^\alpha(\mu_1 \circ \mu_2) = ABS_H^\alpha(\mu_1) \circ ABS_H^\alpha(\mu_2)$$

$$AB4 \quad ABS_H^\alpha(\mu_1 \vee \mu_2) = ABS_H^\alpha(\mu_1) \vee ABS_H^\alpha(\mu_2)$$

$$AB5 \quad ABS_H^\alpha(\mu_1 \wedge \mu_2) = ABS_H^\alpha(\mu_1) \wedge ABS_H^\alpha(\mu_2)$$

Операция унификации базируется на четырех основных правилах:

$$UN1 \quad \begin{aligned} F_i = \varphi_1 \circ F_j \\ F_j = \varphi_1 \circ F_k \end{aligned} \equiv F_N = \varphi_1 \circ F_k, \text{ где } N = \{i, j\}$$

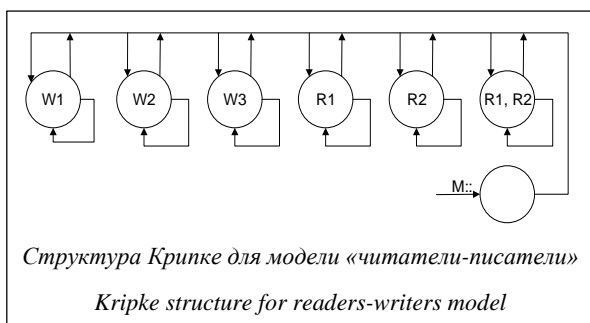
$$UN2 \quad F_N = \varphi \circ F_k \equiv F_N = \varphi \circ F_N, \text{ где } N = \{i_1, i_2, \dots, i_n\}, k \in N$$

$$UN3 \quad F_N = \varphi \circ F_k \equiv F_N = \varphi \circ F_k, \text{ где } N = \{i_1, i_2, \dots, i_n\}, k \notin N$$

$$UN4 \quad F_m = \varphi \circ F_k \equiv F_m = \varphi \circ F_N, \text{ где } N = \{i_1, i_2, \dots, i_n\}, k \in N, m \notin N$$

Их использование позволяет сократить число состояний системы, а также число ее переходов, где N – множество унифицируемых индексов.

В качестве примера возьмем некоторый вариант модели алгоритма «читатели-писатели» для трех «писателей» и двух «читателей». Модель соответствующей системы представлена в виде структуры Крипке на рисунке.



Сформулируем модель данной системы в терминах RLTL-нотации. Каждому состоянию модели Крипке будет сопоставлена своя метаварiable, определяющая соответствующее уравнение системы. В качестве алфавита будет выступать $\Sigma = \{W_1, W_2, W_3, R_1, R_2, R_{12}\}$. Переходы, ведущие из конкретного состояния, будут помечены символом этого состояния. Поскольку первое состояние си-

стемы не помечено ни одним символом, в RLTL все переходы из данного состояния будут помечены символом неопределенности, или Δ .

$$\begin{aligned} F_M &= \Delta \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{W1} &= W_1 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{W2} &= W_2 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{W3} &= W_3 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{R1} &= R_1 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{R2} &= R_2 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{R12} &= R_{12} \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \end{aligned}$$

Важно учитывать, относительно чего планируется производить упрощение системы. Так, в данном случае пусть проверка системы будет выполняться относительно следующего условия, заданного в терминах LTL: $\phi = GF(W_i)$ – «Всегда в будущем процесс записи для любого из «писателей» будет обязательно завершен». Как видно из формулы, в процессе верификации не будет анализироваться ни один из «читателей», так как в формуле не участвуют предикаты $R1$ и $R2$. В связи с этим можно выполнить абстракцию модели системы относительно всех «читателей».

Применим представленные аксиомы к описанному выше примеру. Поскольку абстракция в этом случае выполняется относительно предикатов $R1$ и $R2$, $H = \{R1, R2, R12\}$, а в качестве символа замены будет использован некоторый символ, например $\alpha = R$. Таким образом, получим:

$$\begin{aligned} \text{ABS}_H^R(F_M) &= F_M = \\ &= \text{ABS}_H^R(\Delta \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12})) \\ \text{ABS}_H^R(F_{W1}) &= F_{W1} = \\ &= \text{ABS}_H^R(W_1 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12})) \\ \text{ABS}_H^R(F_{W2}) &= F_{W2} = \\ &= \text{ABS}_H^R(W_2 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12})) \\ \text{ABS}_H^R(F_{W3}) &= F_{W3} = \\ &= \text{ABS}_H^R(W_3 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12})) \\ \text{ABS}_H^R(F_{R1}) &= F_{R1} = \\ &= \text{ABS}_H^R(R_1 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12})) \\ \text{ABS}_H^R(F_{R2}) &= F_{R2} = \\ &= \text{ABS}_H^R(R_2 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12})) \\ \text{ABS}_H^R(F_{R12}) &= F_{R12} = \\ &= \text{ABS}_H^R(R_{12} \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12})) \end{aligned}$$

После применения аксиом AB1–AB5 система уравнений примет следующий вид:

$$\begin{aligned} F_M &= \Delta \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{W1} &= W_1 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{W2} &= W_2 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{W3} &= W_3 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{R1} &= R \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{R2} &= R \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{R12} &= R \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \end{aligned}$$

Применение представленных выше аксиом позволило сократить общее число символов алфавита (предикатов) системы, однако число ее состояний и переходов осталось прежним. Правило UN1 позволяет сократить число состояний модели верифицируемой системы путем унификации однотипных элементов вычислительного пути. Применим данное правило к системе, полученной на предыдущем шаге, где $N = \{R1, R2, R12\}$:

$$\begin{aligned} F_M &= \Delta \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{W1} &= W_1 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{W2} &= W_2 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_{W3} &= W_3 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \\ F_N &= R \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_{R1} \vee F_{R2} \vee F_{R12}) \end{aligned}$$

Применение правил UN2–UN4 позволяет унифицировать однотипные переходы RLTL-модели. После применения данных правил модель примет вид:

$$\begin{aligned} F_M &= \Delta \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_N \vee F_N \vee F_N) \\ F_{W1} &= W_1 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_N \vee F_N \vee F_N) \\ F_{W2} &= W_2 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_N \vee F_N \vee F_N) \\ F_{W3} &= W_3 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_N \vee F_N \vee F_N) \\ F_N &= R \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_N \vee F_N \vee F_N) \end{aligned}$$

Выполнив правило дизъюнкции, получим упрощенную модель системы «читатели-писатели»:

$$\begin{aligned} F_M &= \Delta \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_N) \\ F_{W1} &= W_1 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_N) \\ F_{W2} &= W_2 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_N) \\ F_{W3} &= W_3 \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_N) \\ F_N &= R \circ (F_{W1} \vee F_{W2} \vee F_{W3} \vee F_N) \end{aligned}$$

Таким образом, на конкретном примере за счет эффективного использования операций абстракции и унификации на RLTL-структурах исходную модель удалось сократить более чем в два раза. Исходная модель – 42 перехода, упрощенная модель – 20 переходов.

Приведем также пример упрощения модели светофора, описанной ранее в терминах RLTL. Пусть проверка этой системы будет выполняться относительно следующего условия, заданного в терминах LTL: $\phi = GF(g)$ – «Всегда в будущем будет загораться зеленый сигнал светофора». Тогда в качестве множества заменяемых символов можно взять $N = \{y, r\}$, а в качестве символа замены используем просто символ s , который будет означать stop, $\alpha = s$. Таким образом, получим:

$$\begin{aligned} \text{ABS}_H^s(F_1) &= \text{ABS}_H^s(g \circ F_2) \\ \text{ABS}_H^s(F_2) &= \text{ABS}_H^s(y \circ F_3) \\ \text{ABS}_H^s(F_3) &= \text{ABS}_H^s(r \circ F_4) \\ \text{ABS}_H^s(F_4) &= \text{ABS}_H^s(y \circ F_1) \end{aligned}$$

После применения аксиом AB1–AB5 система уравнений примет следующий вид:

$$\begin{aligned} F_1 &= g \circ F_2 \\ F_2 &= s \circ F_3 \\ F_3 &= s \circ F_4 \\ F_4 &= s \circ F_1 \end{aligned}$$

Для сокращения числа состояний системы применим к ней правило UN1, где $N = \{2, 3, 4\}$, и получим: $F_1 = g \circ F_2$, $F_N = s \circ F_1$.

И наконец, применив правила UN2–UN4, получим сокращенную модель светофора: $F_1 = g \circ F_N$, $F_N = s \circ F_1$.

В результате после выполнения операций абстракции и унификации модель светофора сократилась ровно в два раза.

В качестве более содержательного примера использования операций абстракции и унификации рассмотрим следующий случай. Пусть требуется выполнить верификацию множества драйверов и/или некоторых подпрограмм операционной системы, которые используют разделяемые ресурсы, например, оборудование или адреса в оперативной памяти, на предмет наличия в их совместной работе тупиков, гонок и взаимных блокировок. В таком случае последовательность шагов будет следующей.

1. Построить модель для каждого драйвера (подпрограммы) на базе RLTL.

2. Построить верифицируемые свойства на базе RLTL.

3. Для каждой модели выполнить операцию абстракции:

- определить множество H как множество символов, соответствующих свойствам системы, отличным от свойств, задействованных при работе с разделяемыми ресурсами;

- в качестве символа замены α взять любой не использованный ранее символ;

- выполнить абстракцию с помощью аксиом AB1–AB5.

4. Для каждой модели выполнить операцию унификации с помощью аксиом UN1–UN4.

5. Выполнить верификацию абстрагированных и унифицированных моделей систем.

За счет выполнения пункта 3 число свойств каждой модели сократится до числа тех свойств, которые сопоставляются операциям взаимодействия, плюс одно свойство, обозначаемое символом α и характеризующее все отличные от операций взаимодействия свойства. Выполнение пункта 4 позволит на порядок и более сократить размеры каждой модели за счет объединения однотипных блоков состояний, полученных после применения

пункта 3. Таким образом, за счет того, что общая доля операций взаимодействия относительно всех операций драйвера крайне мала, в данном примере удастся добиться многократного сокращения размеров верифицируемых моделей.

Заключение

Авторами была рассмотрена RLTL-нотация в качестве новой структуры данных для задания моделей систем, что позволяет повысить эффективность и быстродействие процесса верификации за счет сокращения числа свойств систем, а также числа состояний и переходов. Было показано, что задание моделей систем с помощью RLTL не сложнее, чем на основе структур Крипке или на базе автоматов Бюхи. Модели, заданные на базе RLTL, как и модели на базе других структур данных, могут быть существенно упрощены за счет использования представленных в статье методов абстракции и унификации, что является важным аспектом для процесса верификации больших распределенных программных систем.

Таким образом, использование RLTL-нотации является важным и необходимым шагом, поскольку задание моделей верифицируемых систем и их требований сразу на базе единой унифицированной структуры позволяет сделать процесс верификации более гибким, быстрым и удобным.

Литература

1. Карпов Ю.Г. Model Checking. Верификация параллельных и распределенных программных систем. СПб: БХВ-Петербург, 2010. 552 с.
2. Кораблин Ю.П. Семантика языков распределенного программирования. М.: Изд-во МЭИ, 1996. 102 с.
3. Kroger F., Merz S. Temporal Logic and State Systems. Springer, 2008, 436 p.
4. Manna Z., Pnueli A. The Temporal Logic of Reactive and Concurrent Systems. Springer-Verlag, NY, 1992, 427 p.
5. Шипов А.А., Кораблин Ю.П. Эквационная характеристика формул LTL // Программные продукты и системы. 2015. № 4. С. 175–179.
6. Кларк Э.М., Грамберг О., Пелед Д. Верификация моделей программ. Model Checking; [пер. с англ.]. М.: Изд-во МЦНМО, 2002. 416 с.
7. Holzman G.J. An Analysis of Bitstate Hashing. Proc. 15th Int. Conf. on Protocol Specification, Testing, and Verification, 1998, pp. 301–314.
8. Olderog E.-R., Apt K.R. Fairness in Parallel Programs: The Transformational Approach. ACM Transactions on Programming Languages and Systems, July 1988, vol. 10, no. 3, pp. 420–455.

SYSTEM MODELS CONSTRUCTION BASED ON LTL FORMULA EQUATIONAL CHARACTERISTICS

Yu.P. Korablin¹, Dr.Sc. (Engineering), Professor, y.p.k@mail.ru

A.A. Shipov², Leading Engineer, a-j-a-1@yandex.ru

¹ Russian State Social University, Vilgelm Pik St. 4, Moscow, 129226, Russian Federation

² Moscow Technological University (MIREA), Vernadsky Ave. 78, Moscow, 119454, Russian Federation

Abstract. Verification of software and technical systems has always been and still remains one of the most significant tasks since appearance of first computing devices. Today there are quite a lot of approaches to solve this problem. However, development of such formal verification method as Model Checking helped to solve the problem of verifying systems representation and to unify verification process for software and technical systems. Its main idea is to transform an original system into a unified form. It means that a verification process requires only a model that would most precisely describe system's behavior.

The article considers the possibility of system model construction using RLTL-notations (Recursive Linear Temporal Logic), which are a recursive representation of formulas of linear temporal logic. However, its usage is not limited to this aspect. The advantage of using RLTL for these purposes is that models based on it might be verified in respect to requirements which are also RLTL-based without casting to any another data structure. It will certainly help to simplify and improve the performance of a verification process. Furthermore, the article describes the formal tools, which allow simplifying RLTL-based models in many cases reducing the number of their states and transitions.

Keywords: verification, Model Checking, RLTL equation characteristics, Kripke structure, Buchi automaton, temporal logic formula, LTL, CTL.

References

1. Karpov Yu.G. *Model Checking. Verifikatsiya parallelnykh i raspredelennykh programmnykh sistem* [Model Checking. Parallel and Distributed Software System Verification]. St. Petersburg, BHV-Peterburg Publ., 2010, 552 p.
2. Korablin Yu.P. *Semantika yazykov raspredelennogo programmirovaniya* [Distributed Programming Language Semantics]. Study guide. V.P. Kutepov (Ed.). Moscow, MEI Publ., 1996, 102 p.
3. Kroger F., Merz S. *Temporal Logic and State Systems*. Springer Publ., 2008, 436 p.
4. Manna Z., Pnueli A. *The Temporal Logic of Reactive and Concurrent Systems: Specification*. Springer-Verlag, NY, 1992, 427 p.
5. Shipov A.A., Korablin Yu.P. LTL formula equational characteristics. *Programmnye produkty i sistemy* [Software & Systems]. 2015, no. 4 (112), pp. 175–179.
6. Clarke E.M., Grumberg O., Peled D. *Model Checking*. MIT Press, 1999, 330 p. (Russ.ed.: Moscow, MTsNMO Publ., 2002, 416 p.).
7. Holzman G.J. An Analysis of Bitstate Hashing. *Proc. 15th Int. Conf. Protocol Specification, Testing, and Verification*. 1998, pp. 301–314.
8. Olderog E.-R., Apt K.R. Fairness in Parallel Programs: The Transformational Approach. *ACM Trans. Programming Languages and Systems*. 1988, vol. 10, no. 3, pp. 420–455.

Примеры библиографического описания статьи

1. Кораблин Ю.П., Шипов А.А. Построение моделей систем на базе эквациональной характеристики формул LTL // Программные продукты и системы. 2017. Т. 30. № 1. С. 61–66; DOI: 10.15827/0236-235X.117.061-066.
2. Korablin Yu.P., Shipov A.A. System models construction based on LTL formula equational characteristics. *Programmnye produkty i sistemy* [Software & Systems]. 2017, vol. 30, no. 1, pp. 61–66 (in Russ.); DOI: 10.15827/0236-235X.117.061-066.

УДК 004.413; 519.816; 519.226.3
DOI: 10.15827/0236-235X.117.067-074

Дата подачи статьи: 15.07.16
2017. Т. 30. № 1. С. 67–74

ИСПОЛЬЗОВАНИЕ ФОРМУЛЫ БАЙЕСА ПРИ ОЦЕНИВАНИИ ВЫПОЛНЕНИЯ ПРАКТИК МОДЕЛИ СММИ®

*Г.И. Кожомбердиева, к.т.н., доцент, kgi-liizht@yandex.ru;
Д.П. Бураков, к.т.н., доцент, burakovdmity8@gmail.com;
М.И. Гарина, к.т.н., доцент, migarina@gmail.com
(Петербургский государственный университет путей сообщения,
Московский просп., 9, г. Санкт-Петербург, 190031, Россия)*

Статья посвящена методике экспертного оценивания (на основе объективных свидетельств) степени осуществления практик, обеспечивающих реализацию целей процессных областей модели СММИ®, разработанной в Институте программной инженерии Университета Карнеги–Меллона (SEI). Формирование подобных оценок необходимо для получения вывода об уровне зрелости процессов разработки ПО, достигнутом организацией-разработчиком.

В условиях неопределенности и/или неполноты исходной информации о выполнении практик СММИ® с целью повышения степени доверия к принимаемым экспертами-оценщиками решениям целесообразно использовать инструментарий, применяемый для принятия решений в слабо формализованных предметных областях. Ранее в работах авторов рассматривались два подхода к формированию оценок: методы нечеткой логики и методы многокритериальной классификации.

В настоящей статье предпринимаются попытки сделать процедуру экспертного оценивания еще более простой и гибкой, расширить возможности ее использования, повысить объективность оценки. Предлагается подход, основанный на использовании известной в теории вероятностей теоремы гипотез (формулы Байеса). При этом степень реализации практики СММИ® оценивается через распределение вероятностей на множестве гипотез о том, что степень реализации достигла одного из предопределенных уровней. Под байесовской оценкой степени реализации практики понимается апостериорное распределение вероятностей, пересмотренное и уточненное в ходе оценивания.

Значения условных вероятностей, используемых при вычислении байесовской оценки, показывают, насколько гипотезы об уровне выполнения практики подтверждаются полученными объективными свидетельствами.

Ключевые слова: СММИ®, процессная область, уровни зрелости, уровни возможностей, оценивание, объективные свидетельства, теория принятия решений, формула Байеса, байесовский подход, экспертное оценивание.

Комплексная модель зрелости СММИ® (Capability Maturity Model Integration) – это широко известный подход к совершенствованию технологических процессов разработки и сопровождения программных продуктов и систем, разработанный в SEI [1]. Специализированная модель СММИ-DEV (СММИ® for Development) используется как руководство по улучшению качества процессов организаций-разработчиков ПО и рекомендуется в том числе для самооценки организации. Актуальной версией СММИ-DEV является версия 1.3, появившаяся в ноябре 2010 года [2]. Несмотря на то, что новые версии руководства не выходили почти шесть лет, интерес к нему со стороны разработчиков ПО и руководителей предприятий не уменьшается. Продолжает продвигать эту модель и компания «Kondakov Consulting» (<http://consulting.kondakov.ru>) – первая в России организация, сертифицированная для проведения оценивания организаций согласно модели СММИ®.

Фундаментальным структурным элементом СММИ® является процессная область. Под нею понимается группа взаимосвязанных практик, совместное выполнение которых позволяет организации достичь набора целей, признанных важными для улучшений в этой области. Под процессами в модели СММИ® понимаются работы, которые рассматриваются как выполнение практик, при этом под практикой понимается некоторая деятель-

ность, способствующая достижению связанной с ней цели. Цели разделяются на *общие* (generic goal – GG) и *специфические* (specific goal – SG). Соответственно практики, связанные с общей целью, также называются общими (generic practice – GP), а практики, связанные со специфической целью, – специфическими (specific practice – SP). GG относятся ко всем процессным областям, а SG всегда сформулированы для конкретной процессной области. Для каждой специфической практики в модели определяются типичные рабочие продукты, представляющие собой образцы результатов ее выполнения. В СММИ-DEV определены 22 процессные области [2].

В модели СММИ® вводится понятие уровня зрелости производственных процессов организации, достижение которого оценивается через достижение соответствующего уровня возможностей во всех процессных областях, приписанных к данному уровню зрелости. Достигнутый уровень возможностей процессной области показывает, насколько хорошо организация осуществляет работы, относящиеся к данной процессной области. Достижение каждого уровня возможностей определяется реализацией соответствующих целей и практик.

Требования к проведению оценивания в рамках модели СММИ® сформулированы в документе ARC (Appraisal Requirements for СММИ®) [3]. Согласно

ему, любой метод оценки качества процесса основывается на анализе проверенных экспертами свидетельств о реализации связанных с процессной областью общих и специфических практик – так называемых объективных свидетельств.

Авторы данной статьи считают весьма полезным внедрение модели CMMI® или используемых в ней методик для оценки и самооценки зрелости процессов в отечественных компаниях, занимающихся разработкой ПО (как частной, так и государственной форм собственности). Однако следует отметить, что, поскольку ARC не содержит описания конкретных способов оценивания объективных свидетельств и качества реализации практик, практическое применение данной методологии упирается в неопределенность того, какие методы и алгоритмы следует применять для получения оценок. В статье [4] модель CMMI® была рассмотрена более детально, в ней также рассматривалась возможность использования методов нечеткой логики для вывода уровней выполнения практик на основе анализа имеющихся объективных свидетельств.

В данной статье рассматривается методика оценки уровня выполнения практик, основанная на байесовском подходе. Формула Байеса [5] используется для получения вероятностных оценок истинности гипотез о том, что степень выполнения практик соответствует каждому из уровней, определенных в CMMI®. Предлагаемая методика базируется на подходе к оцениванию качества управленческих решений в железнодорожной отрасли, предложенном в [6] и [7]. В частности, в [7] одним из авторов данной статьи был предложен простой способ определения условных вероятностей, используемых в формуле Байеса, как частот попадания экспертных отметок по интегрированной группе показателей качества в пересекающиеся интервалы 10-балльной метрической шкалы, соответствующие уровням ранжирования качества решения.

Применение формулы Байеса упрощает задачу оценивания по сравнению с методами нечеткой логики: уменьшается доля самовольности ЛПР, неизбежной при определении таких параметров нечеткого логического вывода, как виды и формы функций принадлежности, способ реализации нечетких логических операций и т.д. [8]. Подобное упрощение целесообразно, так как высокая точность при оценивании выполнения практик в любом случае невозможна да и не требуется в силу неточности исходных данных и экспертного способа получения оценок. Здесь уместно вспомнить мнение выдающегося математика, академика АН СССР Н.Н. Моисеева, который в контексте обсуждения экспертиз и неформальных процедур в [9] утверждал, что иногда для нужд практики достаточно использовать весьма грубые оценки. Кроме того, байесовский подход позволяет сгладить разногласия, неизбежно возникающие между экспертами

(даже при условии наличия у каждого из них достаточного уровня профессиональной компетентности, исключающего сильные разногласия в оценивании), и освободить лицо, принимающее решение, от необходимости рассчитывать согласованность оценок группы экспертов [10]. Дополнительным доводом в пользу применения байесовского подхода в новом контексте является то, что он давно и успешно используется при принятии решений в условиях неопределенности: при решении задач организационного управления, в том числе задач управления рисками [11–13], при оценивании качества продукции на основании случайного выборочного контроля.

Байесовское оценивание уровней выполнения практик процессной области

Рассмотрим предлагаемый способ применения байесовского подхода к оцениванию уровней выполнения практик процессных областей CMMI. В модели CMMI степень выполнения каждой практики может достигать одного из пяти уровней, упорядоченных по возрастанию качества реализации. Возможные уровни выполнения практики представлены в таблице 1.

Обозначим через H_i гипотезу (hypothesis) вида «Выбранная практика достигает уровня реализации i », где $i = \overline{1, 5}$ – порядковый номер уровня из таблицы 1. Далее предположим, что в распоряжении лица, принимающего решение, имеется n объективных свидетельств (evidence) за или против каждой из гипотез H_i . В качестве свидетельств могут использоваться документы, представляющие результат реализации практики либо являющиеся следствием ее выполнения, а также устные или письменные заявления, подтверждающие осуществление (или невыполнение) практики, предоставляемые ее исполнителями. Факт наличия каждого из свидетельств обозначим через e_j . Отметим, что в отличие от гипотез свидетельства e_j никак не упорядочены по качеству (значимости) и пронумерованы в произвольном порядке.

1. Перед началом оценивания лицо, принимающее решение, формирует априорное распределение вероятностей $P(H_i)$ на множестве гипотез. Каждая вероятность $P(H_i)$ рассматривается как степень уверенности этого лица в справедливости i -й гипотезы об уровне выполнения практики до начала оценивания, то есть до получения каких-либо свидетельств за или против гипотезы. Так как априорная информация об уровне выполнения практики может быть полностью неопределенной, вероятности $P(H_i)$ могут иметь значение $1/5$, $i = \overline{1, 5}$. При наличии достаточного обоснования допускается использование и неравномерного распределения априорных вероятностей на множестве гипотез. Например, крайние гипотезы H_1 и H_5 представля-

Таблица 1

Уровни выполнения практики СММИ

Table 1

CMMI practice implementation levels

№	Уровень выполнения практики	Описание
1	Not Yet (NY)	Практика отсутствует. Отсутствуют свидетельства, позволяющие признать практику реализованной
2	Not Implemented (NI)	Практика не реализована. Предоставленные свидетельства не позволяют заключить, что практика осуществлена, отмечены явные недостатки
3	Partially Implemented (PI)	Практика частично реализована. Предоставленные свидетельства противоречивы: некоторые данные указывают на выполнение практики, некоторые – на невыполнение. Отмечены недостатки
4	Largely Implemented (LI)	Практика в основном реализована. Имеется достаточно убедительных объективных свидетельств, отмечены отдельные недостатки
5	Fully Implemented (FI)	Практика полностью реализована. Имеется достаточно убедительных объективных свидетельств, отсутствуют недостатки

ются менее вероятными, чем все остальные, поэтому их априорные вероятности могут иметь более низкие значения. Кроме того, в качестве априорных вероятностей $P(H_i)$ могут использоваться апостериорные байесовские вероятности $P(H_i | e_1, e_2, \dots, e_n)$, полученные на предыдущей итерации оценивания.

2. Условная вероятность $P(e_j | H_i)$ понимается как вероятность истинности свидетельства e_j в предположении, что истинна гипотеза H_i , и показывает, насколько данные, полученные из свидетельства, соответствуют i -й гипотезе об уровне выполнения практики. Значение этой условной вероятности получается путем агрегации полученных балльных экспертных оценок имеющегося свидетельства. Назначенные экспертами баллы показывают, насколько, по их мнению, каждая из гипотез подтверждается полученным свидетельством, и отражают степени предпочтения экспертами, производящими оценивание, той или иной гипотезы о достижении определенного уровня реализации рассматриваемой практики.

3. Условная вероятность $P(H_i | e_1, e_2, \dots, e_n)$ понимается как степень уверенности лица, производящего оценивание, в справедливости i -й гипотезы об уровне выполнения практики после получения всех свидетельств e_j , $j = \overline{1, n}$. В соответствии с теоремой Байеса и при условии независимости всех свидетельств она вычисляется как апостериорная байесовская вероятность:

$$P(H_i | e_1, \dots, e_n) = \frac{P(H_i) \cdot P(e_1 | H_i) \cdot P(e_2 | H_i) \cdot \dots \cdot P(e_n | H_i)}{\sum_{j=1}^5 P(H_j) \cdot P(e_1 | H_j) \cdot P(e_2 | H_j) \cdot \dots \cdot P(e_n | H_j)} \quad (1)$$

4. Полученное по формуле (1) апостериорное распределение вероятностей $P(H_i | e_1, e_2, \dots, e_n)$ на множестве гипотез, $i = \overline{1, 5}$, является итоговой

оценкой уровня выполнения практики и показывает, насколько правдоподобными по завершении процедуры оценивания стали гипотезы о том, что степень выполнения рассматриваемой практики достигла каждого из уровней.

Рассмотрим простой способ получения и агрегации экспертных оценок для определения условных вероятностей $P(e_j | H_i)$ путем обработки объективных свидетельств. В каждом конкретном случае набор объективных свидетельств определяется как целями оцениваемой организации и типом разрабатываемых продуктов, так и принятым в организации способом фиксации требований к разработке.

При необходимости каждое объективное свидетельство может быть оценено в ходе нескольких экспертиз с использованием различных экспертов или экспертных групп. Чем больше используется объективных свидетельств и проводится экспертиз (при условии адекватной профессиональной компетентности проводящих их экспертов), тем точнее будет полученная общая оценка уровня выполнения соответствующих практик. Для фиксации результатов экспертизы объективных свидетельств уместно использовать контрольные списки (check-list), широко применяемые при оценивании качества процессов или продукции [14].

Экспертные оценки соответствия свидетельств e_j гипотезам о степени выполнения некоторой практики СММИ $P(e_j | H_i)$ формируются по результатам обработки контрольных списков следующим образом.

1. Результаты k -й экспертизы контрольного списка по свидетельству e_j суммируются, а итоговое значение e_{jk} переводится в 10-балльную шкалу для обеспечения однородности экспертных оценок.

2. Для каждого свидетельства e_j подсчитываются относительные частоты попадания всех итоговых значений e_{jk} в частично пересекающиеся интервалы, определенные на 10-балльной шкале и со-

ответствующие пяти гипотезам из таблицы 1, например, NY : [0; 1], NI : [1; 2], PI : [2; 6], LI : [5; 9], FI : [9; 10]. Пересечение интервалов введено намеренно с целью моделирования неопределенности, возникающей при экспертном оценивании, в частности, в связи с использованием свидетельств разного уровня значимости (качества). Более того, для различных свидетельств степень пересечения интервалов может варьироваться в зависимости от уровня их значимости.

3. Полученные относительные частоты и принимаются за оценки условных вероятностей соответствия свидетельств e_j гипотезам об уровне выполнения практики $P(e_j|H_i)$. Они, разумеется, являются очень грубым приближением к условным вероятностям, но, как упоминалось выше, в случае оперирования весьма неопределенными исходными данными большая точность и не требуется.

Оценивание практик процессной области «Разработка требований»

Назначение процессной области «Разработка требований» (Requirement Development, RD) – выявление, анализ и фиксация требований заказчика, а также технических требований и ко всему продукту, и к его компонентам. Требования касаются как в целом функциональности, безопасности, надежности, модифицируемости и масштабируемости продукта, его интегрируемости с внешними приложениями, так и конкретных принимаемых архитектурных решений и определяют действия всех участников проекта по его разработке.

В процессной области имеются следующие специфические цели SG и связанные с ними практики SP .

1. $SG1$ – *Develop Customer Requirements*. Сбор и перевод в требования заказчика пожеланий всех заинтересованных лиц, их ожиданий, ограничений и представлений об интерфейсах разрабатываемого продукта.

- $SP 1.1$ – *Elicit Needs*. Выявление пожеланий заинтересованных лиц, их ожиданий, ограничений и представлений об интерфейсах разрабатываемого продукта на всех фазах жизненного цикла.

- $SP 1.2$ – *Transform Stakeholders Needs into Customer Requirements*. Преобразование пожеланий заинтересованных лиц, их ожиданий, ограничений и представлений об интерфейсах разрабатываемого продукта в перечень требований заказчика с приоритетами.

Результатами выполнения практик цели $SG1$ могут являться

- перечень требований заказчика с приоритетами;
- порядок проведения верификации;
- порядок проведения валидации и т.д.

2. $SG2$ – *Develop Product Requirements*. Разработка технических требований к продукту и его

компонентам путем совершенствования и уточнения требований заказчика.

- $SP 2.1$ – *Establish Product and Product Component Requirements*. Установление и сохранение технических требований к продукту и его компонентам на основе требований заказчика.

- $SP 2.2$ – *Allocate Product Component Requirements*. Распределение требований по компонентам продукта.

- $SP 2.3$ – *Identify Interface Requirements*. Выявление интерфейсных требований (то есть требований к способам информационного обмена между программными функциями, объектами и другими элементами).

Результатами выполнения практик цели $SG2$ могут являться

- общие требования к продукту;
- требования к компонентам продукта, в том числе таблицы распределения требований по компонентам;
- требования к архитектуре, в том числе к связям между компонентами;
- требования к интерфейсам между элементами компонентов;
- проектные ограничения, в том числе внешние, и т.д.

3. $SG3$ – *Analyze and Validate Requirements*. Анализ и валидация требований.

- $SP 3.1$ – *Establish Operational Concepts and Scenarios*. Установление общей концепции процесса разработки и набора реализующих ее сценариев.

- $SP 3.2$ – *Establish of Definition of Required Functionality and Quality Attributes*. Определение требуемой функциональности и критериев качества.

- $SP 3.3$ – *Analyze Requirements*. Анализ требований с точки зрения выявления их необходимости и достаточности.

- $SP 3.4$ – *Analyze Requirements to Achieved Balance*. Анализ требований с точки зрения поиска компромисса между пожеланиями заинтересованных лиц и выявленными ограничениями.

- $SP 3.5$ – *Validate Requirements*. Анализ и проверка требований для гарантии того, что разрабатываемый продукт будет функционировать корректно в среде конечного пользователя.

Результатами выполнения практик цели $SG3$ могут являться

- общая концепция процесса разработки;
- концепции процессов разработки компонентов, установки продукта, его сопровождения и поддержки;
- сценарии, реализующие общую концепцию процесса;
- требования к функциональности продукта;
- сформулированные критерии качества и технической эффективности;

- варианты использования продукта;
- диаграммы активности для вариантов использования;
- функциональная архитектура (выявленные методы и их взаимодействие);
- результаты объектно-ориентированного анализа функциональной архитектуры;
- отчет о недостатках системы требований и рекомендации по их устранению;
- оценка рисков, связанных с требованиями;
- новые дополнительные требования и ограничения.

Нетрудно заметить, что структура и содержание процессной области *RD* на практике в достаточной степени отражается в документации, сопровождающей разработку ПО, в том числе и в отечественной практике. В частности, многие позиции отражаются в *техническом задании на разработку автоматизированной системы* (ТЗ АС), соответствующем требованиям ГОСТ 34.602-89.

Таким образом, экспертная оценка наполнения соответствующих пунктов ТЗ АС может служить объективным свидетельством выполнения специфической практики *SP 2.2*, например:

- *OE1* – п. 4.1.1.1 «Перечень подсистем, их назначение и основные характеристики, требования к числу уровней иерархии и степени централизации системы»;
- *OE2* – п. 4.1.8 «Требования к эксплуатации, техническому обслуживанию, ремонту и хранению компонентов системы»;
- *OE3* – п. 4.2.1 «Перечень функций, задач или их комплексов, подлежащих автоматизации, для каждой подсистемы».

В качестве других объективных свидетельств, подтверждающих выполнение данной практики, можно использовать, например, результаты экспертного оценивания протокола первого совещания с заказчиком (*OE4*), протокола повторного совещания с заказчиком (*OE5*), а также иных зафиксированных в модели СММИ возможных результатов выполнения практик. Отметим, что номера, присвоенные объективным свидетельствам, никак не характеризуют их важность или приоритет с точки зрения лица, принимающего решение, а служат лишь для их идентификации.

Пример байесовского оценивания практики *SP 2.2* процессной области *RD*

Для вычисления байесовской оценки степени выполнения практики *SP 2.2* используем объективные свидетельства *OE1–OE4*, предложенные выше. Рассмотрим процесс формирования оценки уровня выполнения практики. Сначала проводятся экспертизы объективных свидетельств по контрольным спискам, составленным в соответствии с целями предприятия и типом разрабатываемого продукта. Например, контрольный список для оценки раз-

дела ТЗ «Перечень функций, задач или их комплексов, подлежащих автоматизации, для каждой подсистемы» (как свидетельства *OE3*) может соответствовать приведенному в таблице 2.

Таблица 2

Пример контрольного списка для оценивания свидетельства *OE3*

Table 2

A check list example to estimate *OE3* evidence

Вопрос	Шкала оценивания
1. Имеется ли в ТЗ соответствующий раздел?	0 – не имеется 1 – имеется
2. <i>m</i> Указан ли перечень функций для <i>m</i> -й подсистемы?	0 – перечень отсутствует 1 – перечень имеется
3. <i>m</i> Описаны ли для <i>m</i> -й подсистемы автоматизируемые ею процессы?	0 – не описаны 1 – описаны
4. <i>m</i> Хорошо ли отделены функции <i>m</i> -й подсистемы друг от друга?	0 – описание не позволяет судить об этом 1 – описание свидетельствует о наличии большого количества пересечений функциональности 2 – пересечения есть, но незначительные 3 – функции хорошо отделены друг от друга
5. <i>m</i> Полностью ли реализуют автоматизируемые <i>m</i> -й подсистемой процессы ее функции?	0 – описание не позволяет судить об этом 1 – описание свидетельствует о значительной неполноте функциональности 2 – полнота реализации автоматизируемых процессов достаточно высока 3 – функции полностью реализуют автоматизируемые подсистемой процессы
6. Не дублируют ли функции разных подсистем друг друга?	0 – описание не позволяет судить об этом 1 – есть большое количество пересечений функциональности разных подсистем 2 – примерно половина подсистем имеет пересекающуюся функциональность 3 – пересечений функциональности мало 4 – функции разных подсистем не дублируют друг друга
...	...

В предлагаемом варианте контрольного списка вопросы 2–5 повторяются блоками по *N* вопросов, где *N* – число подсистем, составляющих разрабатываемую АС: $m = \overline{1, N}$.

Обработка результатов экспертиз (то есть заполненных контрольных списков) производится в

Таблица 3

Пример расчета условных вероятностей соответствия свидетельств гипотезам

Table 3

An example of calculation conditional probabilities of correspondence between evidence and hypotheses

Приведенные экспертные оценки	NI				LI						
	NY		PI						FI		
	0	1	2	3	4	5	6	7	8	9	10
OE1 – п. 4.1.1.1 Т3 AC											
Э1							*				
Э2									*		
Э3										*	
Э4					*						
Э5			*								
$P(e_1 H_i)$	–	1/5		–		3/5				–	
	0/5		3/5				–		1/5		
OE2 – п. 4.1.8 Т3 AC											
Э1						*					
Э2								*			
Э3										*	
Э4			*								
Э5						*					
$P(e_2 H_i)$	–	1/5		–		4/5				–	
	0/5		3/5				–		1/5		
OE3 – п. 4.2.1 Т3 AC											
Э1						*					
Э2			*								
Э3									*		
Э4							*				
Э5										*	
$P(e_3 H_i)$	–	1/5		–		4/5				–	
	0/5		3/5				–		1/5		
OE4 – протокол первого совещания с заказчиком											
Э1		*									
Э2				*							
Э3					*						
Э4										*	
Э5				*							
$P(e_4 H_i)$	–	1/5		–		1/5				–	
	1/5		3/5				–		1/5		

соответствии с алгоритмом вычисления частотных оценок условных вероятностей $P(e_j|H_i)$, рассмотренным выше. Приведем пример расчета условных вероятностей $P(e_j|H_i)$ и итоговых апостериорных байесовских оценок $P(H_i|e_1, e_2, \dots, e_n)$ уровней выполнения специфической практики SP 2.2 с использованием формулы Байеса. В таблице 3 показан пример расчета условных вероятностей $P(e_j|H_i)$, $i = 1, 5$, $j = 1, 4$, на основании агрегирования преобразованных в 10-балльную шкалу оценок пяти экспертов, полученных по результатам обработки заполненных ими контрольных списков. Результаты вычисления апостериорных байесовских оценок уровней выполнения практики SP 2.2 приведены в таблице 4.

Полученная в примере апостериорная вероятность гипотезы о том, что по результатам обследования практика SP 2.2 достигла третьего уровня реализации (PI, «частично выполнена»), гораздо

выше, чем вероятность истинности прочих гипотез. Следующей по величине апостериорной вероятности является гипотеза о выполнении практики «в основном» (LI), вероятности же прочих гипотез либо нулевые, либо почти равны нулю. Так как предложенный метод рекомендуется в основном для самооценки предприятия, заключением по результатам данного оценивания может быть решение о том, что практику SP 2.2 можно считать реализованной.

Заключение

В статье предложено применение байесовского подхода к оцениванию уровня выполнения практик, определенных в модели СММИ®. Данный подход предполагает использование формулы Байеса для построения распределения апостериорных вероятностей на множестве гипотез о том, что сте-

Таблица 4

Пример байесовской оценки уровня выполнения практики

Table 4

An example of Bayesian estimate of a practice implementation level

Уровень выполнения практики		<i>NY</i>	<i>NI</i>	<i>PI</i>	<i>LI</i>	<i>FI</i>
Номер <i>i</i> -й гипотезы об уровне выполнения		1	2	3	4	5
Априорные вероятности $P(H_i)$		1/5	1/5	1/5	1/5	1/5
Оценки соответствия гипотезам	$P(e_1 H_i)$	0	1/5	3/5	3/5	1/5
	$P(e_2 H_i)$	0	1/5	3/5	4/5	1/5
	$P(e_3 H_i)$	0	1/5	3/5	4/5	1/5
	$P(e_4 H_i)$	1/5	1/5	3/5	1/5	1/5
Апостериорные вероятности $P(H_i e_1, e_2, e_3, e_4)$		0	0,01	0,62	0,36	0,01

пень реализации рассматриваемой практики достигла некоторого возможного уровня, исходя из полученных результатов экспертного оценивания имеющихся объективных свидетельств. Подход не накладывает никаких ограничений на количество, качество и конкретный перечень используемых свидетельств, а также на состав оценивающей экспертной группы и применим не только для специфических практик процессных областей, но и для общих практик при условии, что имеются объективные свидетельства, позволяющие произвести экспертное оценивание. Для унификации результатов экспертизы предлагается использование контрольных списков, а для упрощения агрегации полученных оценок и их пересчета в условные вероятности гипотез – перевод всех оценок в единую (например 10-балльную) шкалу.

Используемые при вычислении байесовской оценки уровня выполнения практики условные вероятности рассматриваются как экспертные оценки соответствия объективных свидетельств гипотезам о достижении того или иного уровня выполнения практики. Они показывают, насколько гипотезы подтверждаются полученными свидетельствами, отражают степень предпочтения, отдаваемого экспертами той или иной гипотезе. При этом точность представления этих оценок не является существенной.

Преимущества предлагаемого байесовского подхода:

- упрощение процедуры оценивания по сравнению с использованием методов нечеткой логики;
- вероятностный, более объективный, характер экспертных оценок уровня выполнения практик, а также естественное сглаживание разногласий, возникающих между экспертами;
- возможность оценивания уровня выполнения практик по ограниченному набору имеющихся объективных свидетельств (и/или экспертной группой ограниченного состава) и получения при этом вполне состоятельных оценок.

Байесовский подход находит применение в менеджменте, при решении задач управления рисками и организации выборочного контроля качества продукции. Этот подход может быть ис-

пользован также для оценивания качества управленческих решений и, по мнению авторов, для выполнения процедур самообследования предприятий в соответствии с критериями, предлагаемыми в модели CMMI®.

Авторы глубоко благодарны крупному специалисту в области проблем управления на железнодорожном транспорте профессору А.Е. Красковскому, поддержавшему идею применения байесовского подхода при оценивании качества управленческих решений, а также выражают искреннюю признательность председателю Совета директоров группы компаний Digital Design А.Р. Фёдорову и бывшему начальнику Департамента информатизации и корпоративных процессов управления ОАО «РЖД» А.В. Илларионову, благодаря которым несколько лет назад открыли для себя модель CMMI®.

Литература

1. Ахен Д.М., Клауз А., Тернер Р. CMMI®: Комплексный подход к совершенствованию процессов. Практическое введение в модель; [пер. с англ.]. М.: МФК, 2005. 330 с.
2. CMMI® for Development (CMMI-DEV, V1.3). Improving processes for developing better products and services. CMU/SEI-2010-TR-033. ESC-TR-2010-033. Software Eng. Inst. CMMI Product Team, November 2010.
3. Appraisal Requirements for CMMI (ARC, V1.2). CMU/SEI-2006-TR-011. ESC-TR-2006-011. Software Eng. Inst. SCAMPI Upgrade Team, August 2006.
4. Кожомбердиева Г.И., Гарина М.И., Бураков Д.П. Об использовании аппарата теории принятия решений в задачах оценивания согласно модели CMMI® // Программные продукты и системы. 2013. № 4. С. 117–124.
5. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения. М.: Наука, 1988. 480 с.
6. Кожомбердиева Г.И., Красковский А.Е. Байесовский подход к оценке качества управленческих решений // Интеллектуальные системы на транспорте: матер. III Междунар. науч.-практич. конф. «ИнтеллектТранс-2013». СПб: Изд-во ПГУПС, 2013. С. 384–391.
7. Кожомбердиева Г.И., Красковский А.Е. Способ определения условных вероятностей при байесовском оценивании качества управленческих решений на железнодорожном транспорте // Интеллектуальные системы на транспорте: матер. IV Междунар. науч.-практич. конф. «ИнтеллектТранс-2014». СПб: Изд-во ПГУПС, 2014. С. 412–418.
8. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. СПб: БХВ-Петербург, 2005. 736 с.
9. Моисеев Н.Н. Математические задачи системного анализа: учеб. пособие. М.: ЛИБРОКОМ, 2012. 488 с.
10. Кендэл М. Ранговые корреляции. Серия: Зарубежные статистические исследования; [пер. с англ.]. М.: Статистика, 1975. 216 с.

11. Райфа Г. Анализ решений (введение в проблему выбора в условиях неопределенности); [пер. с англ.]. М.: Наука, 1977. 408 с.

12. Моррис У.Т. Наука об управлении: байесовский подход; [пер. с англ.]. М.: Мир, 1971. 304 с.

13. Уткин Л. В. Анализ риска и принятие решений при неполной информации. СПб: Наука, 2007. 404 с.

14. Scriven M. The logic and methodology of checklists. URL: http://www.wmich.edu/sites/default/files/attachments/u350/2014/logic&methodology_dec07.pdf (дата обращения: 01.07.2016).

Software & Systems

DOI: 10.15827/0236-235X.117.067-074

Received 15.07.16

2017, vol. 30, no. 1, pp. 67–74

USING BAYES' THEOREM TO ESTIMATE CMMI® PRACTICES IMPLEMENTATION

G.I. Kozhombierdieva¹, Ph.D. (Engineering), Associate Professor, kgi-liizht@yandex.ru

D.P. Burakov¹, Ph.D. (Engineering), Associate Professor, burakovdmityr8@gmail.com

M.I. Garina¹, Ph.D. (Engineering), Associate Professor, migarina@gmail.com

¹ Petersburg State Transport University, Moskovsky Ave. 9, St. Petersburg, 190031, Russian Federation

Abstract. The article is devoted to the expert estimation methodology (based on objective evidence) for appraising the extent of implementation of practices, which ensure achievement of the goals of CMMI® model process areas. The model has been developed by the Software Engineering Institute (SEI) at Carnegie Mellon University. Such appraisals are necessary to understand the software development processes maturity level in a developer company.

In case of uncertainty and/or incompleteness of information on CMMI® practice implementation, it is reasonable to use a toolkit for decision-making in weakly formalized subject domains. It helps to increase a degree of belief to decisions of appraisal team members. In the previously published work, the authors have considered two approaches to construction of the estimation: fuzzy logic methods and multi-criteria classification methods.

This article makes an attempt to make the appraisal procedure even more simple and flexible, to expand the opportunities for its use and to increase its objectivity. The proposed approach is based on the known Bayes' theorem. An extent of CMMI® practice implementation is estimated via the distribution of probabilities on a set of hypotheses. Each of hypotheses assumes that an implementation level reached one of predefined ones. The Bayesian estimation of a practice implementation extent is understood as a posteriori probability distribution, which is revised and refined during the estimation. Values of conditional probability that are used when calculating the Bayesian estimation, show how much hypothesis on a practices implementation level are supported by the obtained objective evidences.

Keywords: CMMI®, process area, capability levels, maturity levels, appraisal, objective evidence, decision theory, Bayes' theorem, Bayesian approach, expert estimation.

Acknowledgements. The authors appreciate cooperation of Prof. A.E. Kraskovsky, who is a qualified specialist in the field of control issues on railway transport. He supports the idea of using Bayes' theorem when estimating the quality of management decisions. We are also grateful to a chairman of the board of directors of Digital Design corporate group A.R. Fedorov and a former head of the Department of IT development and corporate processes of Russian Railways A.V. Illarionov, who helped to discover CMMI® model.

References

1. Ahern D.M., Clouse A., Turner R. *CMMI Distilled: A Practical Introduction to Integrated Process Improvement*. Addison-Wesley Prof. Publ., 2004, 305 p. (Russ.ed.: Moscow, MFK Publ., 2005, 330 p.).
2. *CMMI® for Development, Version 1.3. (CMMI-DEV, V1.3)*. Improving processes for developing better products and services. CMU/SEI-2010-TR-033. ESC-TR-2010-033. Software Engineering Institute, CMMI Product Team, 2010.
3. *Appraisal Requirements for CMMI®, Version 1.2 (ARC, V1.2)*. CMU/SEI-2006-TR-011. ESC-TR-2006-011. Software Engineering Institute, SCAMPI Upgrade Team, 2006.
4. Kozhombierdieva G.I., Garina M.I., Burakov D.P. Using decision making theory for appraisal problems according to the CMMI® model. *Programmnye produkty i sistemy* [Software & Systems]. 2013, no. 4, pp. 117–124 (in Russ.).
5. Venttsel E.S., Ovcharov L.A. *Teoriya veroyatnostey i ee inzhenernye prilozheniya* [Theory of Probability and its Engineering Applications]. Moscow, Nauka Publ., 1988, 480 p.
6. Kozhombierdieva G.I., Kraskovsky A.E. Bayesian approach to control decision quality estimation. *Intellektualnye sistemy na transporte: mater. III mezhdunar. nauch.-praktich. konf. "IntellectTrans-2013"* [Intelligent Systems on Transport: Proc. 3rd int. Science and Practice Conf. "IntellectTrans-2013"]. St. Petersburg State Transport Univ. Publ., 2013, pp. 384–391 (in Russ.).
7. Kozhombierdieva G.I., Kraskovsky A.E. A method of conditional probabilities determination for Bayesian control decision quality estimation on transport. *Intellektualnye sistemy na transporte: mater. IV mezhdunar. nauch.-praktich. konf. "IntellectTrans-2014"* [Intelligent Systems on Transport: Proc. 4th Int. Science and Practice. Conf. "IntellectTrans-2014"]. St. Petersburg State Transport Univ. Publ., 2014, pp. 412–418 (in Russ.).
8. Leonenkov A.V. *Nechetkoe modelirovanie v srede MATLAB i fuzzyTECH* [Fuzzy Modeling in MATLAB and fuzzyTECH]. St. Petersburg, BHV-Peterburg Publ., 2005, 736 p.
9. Moiseev N.N. *Matematicheskie zadachi sistemnogo analiza* [Math Tasks of System Analysis]. Study guide. 2nd ed. Moscow, LIBROKOM Publ., 2012, 488 p.
10. Kendall M. *Rank Correlation Methods*. 4th ed. Griffin, London, 1970.
11. Raiffa H. *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*. Addison-Wesley Publ., Reading, MA, 1968.
12. Morris W. *Management science: A Bayesian introduction*. Prentice-Hall Publ., Englewood Cliffs, NY, 1968.
13. Utkin L.V. *Analiz riska i prinyatie resheny pri nepolnoy informatsii* [Risk analysis and decision making with incomplete information]. St. Petersburg, Nauka Publ., 2007, 404 p. (in Russ.).
14. Scriven M. *The logic and methodology of checklists*. 2000. Available at: http://www.wmich.edu/sites/default/files/attachments/u350/2014/logic&methodology_dec07.pdf (accessed July 1, 2016).

УДК 519.677

DOI: 10.15827/0236-235X.117.075-080

Дата подачи статьи: 27.07.16

2017. Т. 30. № 1. С. 75–80

МНОГОПРОЦЕССОРНАЯ ОБРАБОТКА В ЗАДАЧЕ ПРОСТРАНСТВЕННОЙ РЕКОНСТРУКЦИИ ПО МНОЖЕСТВУ ВИДОВ

*В.А. Бобков, д.т.н., зав. лабораторией, bobkov@iacp.dvo.ru;
А.П. Кудряшов, к.т.н., младший научный сотрудник, kudryashova@dvo.ru;
С.В. Мельман, к.т.н., младший научный сотрудник, melman@dvo.ru
(Институт автоматики и процессов управления ДВО РАН,
ул. Радио, 5, г. Владивосток, 690041, Россия)*

Предлагается схема многопроцессорной обработки больших объемов пространственных данных на базе гибридного вычислительного кластера применительно к воксельному методу построения и визуализации 3D-модели сцены подводной обстановки. Рассматривается вычислительная схема воксельного метода, которая состоит из нескольких этапов обработки данных, включая загрузку исходных карт глубин многих видов, построение воксельного представления скалярного поля и построение изоповерхности по воксельному пространству. Вычислительная схема анализируется с точки зрения выявления наиболее вычислительно трудоемких этапов работы и целесообразности организации многопроцессорной обработки.

Также рассматривается архитектура используемого гибридного вычислительного кластера, объединяющая три уровня многопроцессорной обработки: вычислительные узлы кластера, многоядерность и графические процессоры видеоплаты. Используются два типа параллельных архитектур: MPI (параллелизм в рамках кластера) и CUDA (параллелизм на графическом ускорителе). Предложенное решение по распределению вычислительной нагрузки основывается на учете характера вычислений на каждом этапе и особенностях используемых параллельных архитектур. Приводится обоснование реализуемой схемы многопроцессорной обработки с качественными и количественными оценками. Реализованная схема обработки данных обеспечивает максимальное ускорение процесса счета применительно к решению задачи 3D-реконструкции сцены на базе рассмотренного вычислительного кластера.

Приведены результаты вычислительных экспериментов с реальными данными, полученными со сканера RangeVisionPremium5 Mrix. Анализ результатов тестирования подтвердил возможность принципиального повышения вычислительной производительности в рассматриваемой задаче за счет организации распределенно-параллельной обработки данных. Аналогичная схема может применяться и в других задачах, связанных с обработкой больших объемов пространственных данных.

Ключевые слова: 3D-реконструкция, гибридный параллелизм, воксельный метод.

Решение ряда важных практических задач, связанных с обследованием морского дна (поиск различных аномальных зон и источников экологического загрязнения акваторий, инспекция подводных коммуникаций, поиск затонувших объектов), требует решения задачи моделирования подводной обстановки по сенсорным данным, получаемым с помощью подводных аппаратов. Одним из наиболее приемлемых источников данных для построения 3D-моделей являются видеоизображения, из которых с помощью методов компьютерного зрения можно получать карты глубин, то есть дальностные данные, связанные с конкретным видом (проекцией) наблюдаемой сцены.

Первые работы основывались на анализе данных в 3D-пространстве сцены и в 2D-пространствах изображений видов. Подобная схема решения была реализована в [1]. Данная схема не исключала дублирования треугольников при объединении множества видов. Более эффективным методом построения 3D-модели по отдельным видам является воксельный метод, впервые предложенный в работе [2]. Данный метод при высоком качестве результата демонстрировал высокую ресурсоемкость, и реконструкция даже небольшой сцены занимает много часов [3]. В работе [4] представлен обзор различных методов реконструкции

поверхности по набору дальностных данных. Воксельный метод построения единой сеточной модели был применен в работе [5] для реконструкции и визуализации протяженных подводных сцен. Обеспечивая хорошее качество получаемой модели, он, однако, требует слишком много вычислительных ресурсов при использовании воксельного пространства с высоким разрешением. Это особенно неприемлемо в случае протяженных подводных сцен. Поэтому во многих работах усилия разработчиков были направлены на повышение алгоритмической скорости обработки данных в этом методе. Например, в работе [6] была предложена алгоритмическая база для повышения скорости обработки за счет как оптимизации вычислений, так и применения технологии CUDA. Также стоит отметить современный сурфелльный подход, где для ускорения используется GPU [7] для стереорекострукции и визуализации, который при высокой скорости и качестве визуализации не дает единого сеточного представления. Использование GPU для ускорения вычислений происходит достаточно часто для решения задачи реконструкции [8], однако принципиальным подходом к ускорению вычислений в задачах обработки больших объемов пространственных данных является использование суперЭВМ с экстремальным параллелизмом. По-

этому в данной работе применительно к задаче построения и визуализации 3D-модели подводных сцен воксельным методом предлагается подход, основанный на реализации распределенно-параллельной обработки данных на базе суперЭВМ.

Параллельная схема алгоритма

Задача формулируется следующим образом. Имеются несколько видов (ракурсов сцены), для каждого из которых на предварительном этапе была построена поверхность в виде пространственной триангуляционной сетки. Триангуляционная сетка строилась по 2D-триангуляции на регулярной пиксельной решетке изображения с использованием карты глубин. Заметим, что карта глубин вычисляется для точек, наблюдаемых на данном виде. Цель – построить единую текстурированную триангуляционную оболочку, которая и будет трехмерной триангуляционной моделью сцены.

Метод работает в воксельном пространстве сцены. Используется непрерывная неявная функция, представленная значениями в узлах воксельной решетки. Более подробное описание алгоритма построения трехмерной модели сцены по множеству ее видов можно найти в [6].

Для построения эффективной схемы распараллеливания алгоритма необходимо выявить наиболее вычислительно трудоемкие этапы его работы. В вычислительном отношении в алгоритме можно выделить следующие этапы:

- загрузка видов в оперативную память;
- определение границ сцены;
- построение индексной карты для каждого вида;
- построение скалярного поля, заданного воксельным представлением;
- построение изоповерхности по воксельному пространству.

Размер всей структуры входных данных зависит от детализации исходных видов и от их количества. Во время загрузки каждой триангуляционной сетки просматриваются координаты их вершин для построения границ всей сцены. Объединяя первый и второй этапы, можно избавиться от вторичного прохода по всем данным и тем самым повысить общую скорость работы схемы. На этапе загрузки основные затраты времени приходятся на работу внешних устройств хранения информации и использование графических ускорителей здесь не оправданно.

Предполагается, что для каждого из видов определены параметры камеры. Третий этап – построение индексной карты – выполняется для каждого вида и заключается в вычислении проекции всех треугольников сетки на экранную плоскость камеры, с которой был получен данный вид. Проекции присваивается номер соответствующего тре-

угольника (индекс). Вычислительная сложность этого этапа определяется следующими параметрами:

- размер триангуляционной сетки (количество треугольников);
- размер/разрешение проекции (задается пользователем и, как правило, выбирается равным размерам исходных видов).

Размер проекции определяется степенью детализации индексной карты, что, в конечном счете, влияет на качество итоговой единой триангуляционной поверхности. Схема построения индексной карты на CUDA имеет одно узкое место. При заполнении индексной карты два треугольника со смежным ребром могут обращаться к одной и той же памяти для внесения своего индекса. В этом случае индекс, который записывается в индексную карту, определяется наименьшим расстоянием точки прообраза до камеры. Такая проверка влияет на эффективность работы алгоритма на CUDA, поскольку возникает взаимозависимость между параллельно вычисляемыми потоками. Каждый поток вычисляет одну пространственную точку фрагмента, и может возникнуть ситуация, когда сравниваются расстояния до камеры для нескольких точек из параллельных потоков. В этом случае целесообразно применять атомарные функции, которые выполняют часть операций последовательно. Тестирование показало, что и в этом случае реализация на CUDA обеспечивает преимущество в производительности в сравнении с CPU-версией.

На четвертом этапе выполняется основная часть вычислений, связанная с построением воксельной модели объекта. Вычислительная сложность на этом этапе определяется только детализацией воксельного представления. Для каждого вокселя необходимо построить проекции этого вокселя на все индексные карты и вычислить значение скалярной. При этом вычисления для каждого вокселя не зависят от других узлов воксельной решетки, и данный метод может быть эффективно реализован на CUDA.

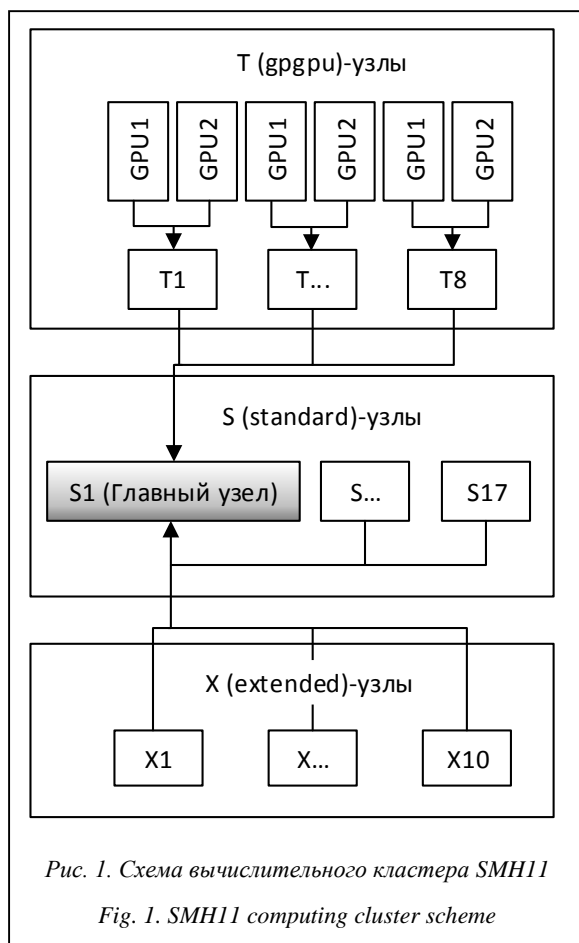
На последнем этапе выполняется построение результирующей изоповерхности в воксельном пространстве методом марширующих кубиков [9]. Здесь была использована известная эффективная реализация на CUDA [10].

Используемый вычислительный кластер

Для решения поставленной задачи использовался гибридный вычислительный кластер Института автоматики и процессов управления ДВО РАН: smh11.cc.dvo.ru. Он представляет собой вычислительную систему из нескольких групп узлов, различающихся аппаратными ресурсами (рис. 1).

1. Простые узлы (17 штук) для MPI-вычислений (Standard):

- процессор: 4xCPU AMD 12-Core Opteron 6164HE;
- частота процессора: 1,7 GHz;
- кэш L1: 12 x 64 KB instruction caches + 12 x 64 KB data caches;
- кэш L2: 12 x 512 KB;
- кэш L3: 2 x 6 MB;
- FSB: 3200 MHz (Hyper Transport links);
- память: 64 Gb оперативной памяти DDR3-1333 MHz ECCReg;
- диск: SSD SATA 2.5" 96 Gb MLC Chip;
- управляющая сеть: Ethernet 1Gb;
- MPI сеть: Infiniband 4xQDR (40 Gb/s) + Ethernet 1Gb.



2. Узлы с увеличенной оперативной памятью (Extended, 10 штук):

- процессор: 4xCPU AMD 12-Core Opteron 6164HE;
- частота процессора: 1,7 GHz;
- кэш L1: 12 x 64 KB instruction caches + 12 x 64 KB data caches;
- кэш L2: 12 x 512 KB;
- кэш L3: 2 x 6 MB;
- FSB: 3200 MHz (Hyper Transport links);
- память: 128 Gb оперативной памяти DDR3-1333 MHz ECC Reg;

- диск: 4xSSD SATA 2.5" 96 Gb MLC Chip.
3. Гибридные узлы (gpgrpu, 8 штук):
- процессор: 2xCPU Xeon L5506 4.8 GTsec;
 - частота процессора: 2.13 GHz;
 - кэш L1: 64 KB;
 - кэш L2: 1024 KB;
 - кэш L3: 4 MB;
 - FSB: 4.8 GT/s QPI (2400 MHz);
 - графический чип: 2xNVIDIA Tesla M2050 GPU computing processor – 3 GB;
 - память: 32 Gb оперативной памяти DDR3-1333MHz ECC Reg;
 - диск: SATA DOM 32 Gb;
 - управляющая сеть: Ethernet 1Gb, MPI сеть: Infiniband 4 QDR (40 Gb/s) + Ethernet 1 Gb.

Коммуникация между узлами возможна только в рамках своей группы. В данной работе использовались узлы типа Т, так как только они имеют по две NVIDIA-видеокарты на узел, соответственно, поддерживают CUDA.

Наличие нескольких узлов позволяет распределять вычислительную нагрузку между ними посредством MPI. Таким образом, используя все узлы, получим 8 CPU-узлов и 16 CUDA-устройств (по 2 на каждый узел). Отметим также, что обмен между CUDA-устройством и хостом возможен только в пределах одного узла.

Распределение вычислительной нагрузки

Эффективность параллельной вычислительной схемы – это эффективное использование имеющихся вычислительных ресурсов. Основными критериями эффективности реализации можно считать максимальное задействование всех доступных ресурсов в каждый момент времени, минимизацию простоев вычислительных узлов и задержек при передаче данных. В работе были использованы два типа параллельных архитектур: MPI (параллелизм в рамках кластера) и CUDA (параллелизм на графическом ускорителе). Только учитывая особенности каждой из архитектур вычислительного комплекса, можно построить эффективную схему распараллеливания предложенного алгоритма.

Преимущества MPI: высокая производительность каждого отдельно взятого потока, высокая скорость передачи данных между потоками. Главный недостаток MPI: малое количество полноценных потоков, определяемое суммарным количеством ядер во всех процессорах кластера. Преимущества CUDA: большое количество параллельно выполняемых потоков. Недостатки CUDA: сравнительно невысокая вычислительная производительность каждого отдельно взятого потока, неэффективный обмен данными между потоками, а также между хостом и устройством, ограниченный объем памяти GPU.

С учетом особенностей гибридного кластера и наличия вычислителей двух видов был обоснован выбор вида параллелизма на каждом этапе вычислительной схемы.

На этапе загрузки видов выполняются считывание с диска и загрузка данных в оперативную память. В данном случае наиболее эффективный способ ускорения процесса – распределение загрузки файлов между узлами MPI. Тогда каждый узел не будет загружать полный набор файлов.

Этап загрузки объединен с этапом поиска границ части объекта, а размер всей области может быть получен с помощью обмена локальными размерами с каждого узла с помощью MPI.

На этапах создания индексных карт, вычисления скалярной функции и построения изоповерхности выполняется большой объем однотипных, независимых вычислений в воксельном пространстве. Поэтому они могут быть эффективно реализованы на CUDA. Тем более что процесс создания индексных карт по сути является процессом растеризации полигональных моделей, то есть естественным для работы графического ускорителя.

Как уже упоминалось, одним из узких мест графических ускорителей является высокая стоимость операции передачи данных между хостом и устройством. Так как все три этапа реализуются на CUDA, в целях минимизации расходов передача данных происходит только дважды: из хоста на устройство копируются параметры воксельного пространства, индексные карты и полигональные модели, а из устройства на хост копируется результирующая изоповерхность.

Кластер содержит 8 узлов с 2 графическими ускорителями на каждом, что позволяет одновременно рассчитывать 16 сегментов объекта. Сегмент – это часть воксельной структуры. Например, воксельная структура размером $1024 \times 1024 \times 1024$ разбивается на 16 сегментов размером $64 \times 1024 \times 1024$. Такой размер сегмента в оперативной памяти графического ускорителя занимает 256 Мб. Если для большей детализации выбрано разбиение по стороне более чем 2^{12} , то размер сегмента в оперативной памяти будет занимать уже более 16 Гб, что значительно больше оперативной памяти графических вычислителей. В этом случае каждый сегмент необходимо еще разбить на более мелкие фрагменты, что не изменит вычислительную схему. А так как готовый полигональный фрагмент всей сцены формируется полностью на GPU, разбиение сегментов на более мелкие фрагменты не повлечет за собой дополнительные расходы на передачу данных между CPU и GPU.

Пусть n – степень разбиения воксельной модели, тогда для хранения данных с плавающей запятой одинарной точности всей воксельной структуры необходимо 2^{3n+2} байт. Если K – количество GPU с памятью, равной M байт, а S – память для хранения исходных данных (индексные карты и ис-

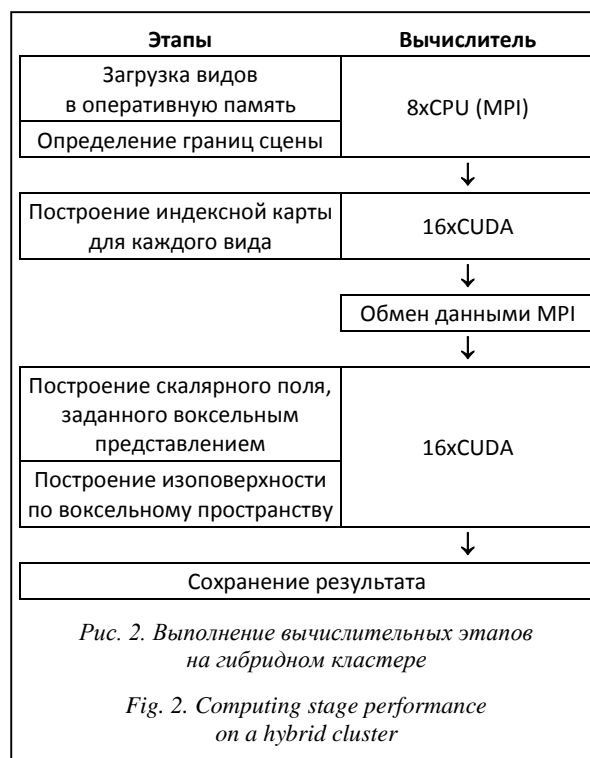
ходные полигональные сетки), то количество фрагментов для каждого узла

$$F = \text{Целая часть} \left(1 + \frac{2^{3n+2}}{(M - S)K} \right).$$

В случае, когда $F > 1$, процесс обработки фрагментов на GPU происходит последовательно.

Применительно к используемому в данной работе гибриднему кластеру для сеток с глубиной менее 11 достаточно 1 фрагмента на каждый GPU. Для $n=12$ уже потребуется 6 фрагментов на GPU.

Поэтапное распределение вычислительной нагрузки на гибридном кластере показано на рисунке 2.



Результаты тестирования

Для получения оценки эффективности предложенной параллельной схемы алгоритма были проведены вычислительные эксперименты на реальной сцене «Пизанская башня» (рис. 3, 4), данные для которой были получены со сканера Range Vision Premium 5 Mpix. Параметры используемого вычислительного оборудования: процессор Intel Core i5 3.3 ГГц и вычислительный кластер, описанный выше. Время работы алгоритма на разных вычислительных устройствах приведено в таблице. Как и ожидалось, наблюдается почти 70-кратное ускорение работы базового алгоритма с применением CUDA даже на одном ускорителе. Применение всех 16 графических ускорителей кластера дает еще 6-7-кратный прирост производительности. Нелинейность ускорения объясняется здесь тем, что

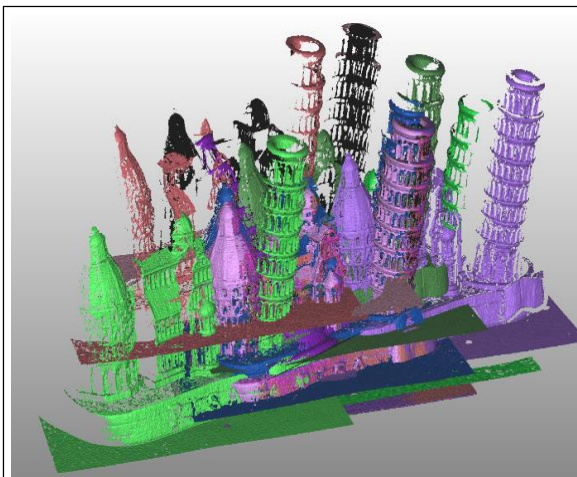


Рис. 3. Исходные виды сцены «Пизанская башня»:
17 видов, 7,2 млн треугольников

Fig. 3. Unmodified state of the scene "Tower of Pisa":
17 views, 7,2 million of triangles

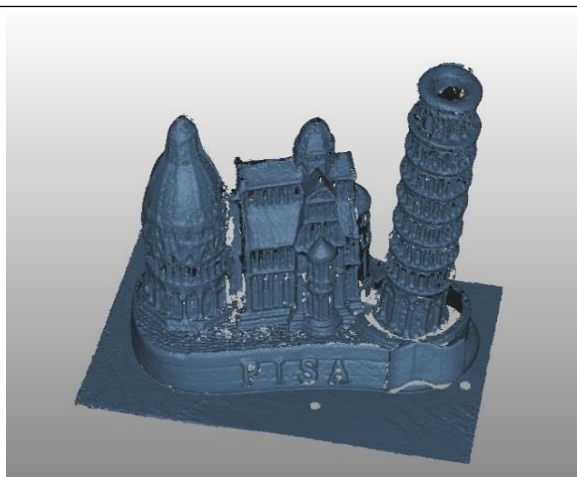


Рис. 4. Реконструкция сцены «Пизанская башня»:
512³ вокселей, 1,2 млн треугольников

Fig. 4. Reconstruction of the scene "Tower of Pisa":
512³ voxels, 1,2 million of triangles

окончательная сборка сцены происходит на одном CPU-кластере.

Результаты тестирования алгоритма на вычислительном кластере для разных разрешений воксельного пространства

The results of algorithm testing on a computing cluster for different voxel resolutions

Вычислитель	Размер сетки			
	512 ³	1024 ³	2048 ³	4096 ³
CPU i5, 3.3 ГГц, 24 Гб ОЗУ, 4 ядра (сек.)	12	75	369	-
CUDA 1xTesla M2050 (сек.)	-	< 1	4,2	67
CUDA 16xTesla M2050 (сек.)	-	< 1	< 1	10,4
Количество треугольников (млн)	1,2	14,8	89,3	447

Заключение

Предложенная схема многопроцессорной обработки на базе гибридного вычислительного кластера в задаче построения и визуализации трехмерной модели сложной сцены по дальностным данным позволила повысить вычислительную производительность более чем на два порядка. Аналогичная схема может применяться и в других задачах, связанных с обработкой больших объемов пространственных данных.

Работа выполнена при финансовой поддержке РФФИ (проект № 15-07-00341), Программ Президиума РАН (проект № 0262-2014-0157) и «Дальний Восток» (проект № 15-I-4-025).

Литература

1. Mordohai P., Frahm J.-M., Akbarzadeh A., etc. Real-time video-based reconstruction of urban. ISPRS Working Group V4 Workshop 3D-ARCH 2007: 3D Virtual Reconstruction and Visualization of Complex Architectures (ETH Zurich, Switzerland). 2007, pp. 121–131.
2. Curless B., Levoy M. A Volumetric method for building complex models from range images. Proc. Conf. Computer Graphics (SIGGRAPH '96). 1996, pp. 303–312.
3. Goesele M., Curless B., Seitz S.M. Multi-view stereo revisited. Computer Vision and Pattern Recognition, Proc. IEEE Comp. Society Conf. 2006, vol. 2, pp. 2402–2409.
4. Seitz S., Curless B., Diebel J., Scharstein D., Szeliski R. A Comparison and evaluation of multi-view stereo reconstruction algorithms. Vision and Pattern Recognition. Proc. IEEE Comp. Society Conf. 2006, vol. 1, pp. 519–526.
5. Johnson-Roberson M., Pizarro O., Williams S.B., Mahon I. Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys. Jour. of Field Robotics, Spec. Iss.: Three-Dimensional Mapping, 2010, vol. 27, iss. 1, part 3, pp. 21–51.
6. Бобков В.А., Кудряшов А.П. Воксельный метод построения триангуляционной поверхности по множеству видов // Информатика и системы управления. 2012. № 2. С. 31–38.
7. Ju Yong Chang, Haesol Park, In Kyu Park, etc. GPU-friendly multi-view stereo reconstruction using surfel representation and graph cuts. Comp. Vision and Image Understanding, 2011, vol. 115, iss. 5, pp. 620–634.
8. Mak J., Hess-Flores M., Recker S., etc. GPU-accelerated and efficient multi-view triangulation for scene reconstruction. Proc. IEEE Winter Conf. on Applications of Comp. Vision, Steamboat Springs, CO, 24–26 March 2014, pp. 61–68.
9. Lorensen W.E., Cline H.E. Marching cubes: A high resolution 3D surface construction algorithm. In Proc. Conf. Computer Graphics (SIGGRAPH '87). 1987, vol. 21, pp. 163–169.
10. Dyken C., Ziegler G. GPU-accelerated data expansion for the Marching Cubes algorithm. In Proc. GPU Technology Conf., San Jose (CA), Sep. 23, 2010, pp. 115–123.

**MULTIPROCESSING FOR SPATIAL RECONSTRUCTION BASED
ON MULTIPLE RANGE-SCANS****V.A. Bobkov¹**, Dr.Sc. (Engineering), Head of Laboratory**A.P. Kudryashov¹**, Ph.D. (Engineering), Junior Researcher**S.V. Melman¹**, Ph.D. (Engineering), Junior Researcher, melman@dvo.ru¹ Institute of Automation and Control Processes Far Eastern Branch of RAS,
Radio St. 5, Vladivostok, 690041, Russian Federation

Abstract. The paper proposes a scheme for multiprocessing large volumes of spatial data based on the hybrid computing cluster. This scheme uses the voxel approach for reconstruction and visualization of 3D models of underwater scenes. There are several processing steps including loading various types of initial depth maps, construction of voxel representation of a scalar field and construction of an isosurface using voxel space. The authors analyze a computational scheme to identify the most computationally intensive stages and to understand whether multiprocessing is feasible.

They also consider the hybrid computing cluster architecture, which combines three levels of multiprocessing: computing nodes, multi-core and GPU video cards. Two types of parallel architectures are used: MPI and CUDA (parallel computing on GPU). The proposed solution of processing load distribution is based on the nature of each stage and the features of used parallel architectures. The paper provides substantiation for the implemented scheme with qualitative and quantitative assessment. The implemented data processing scheme provides a maximum acceleration of a scene 3D reconstruction using the considered computational cluster.

The paper presents the results of computational experiments with real data obtained from the scanner RangeVision Premium 5 Mpix. Test result analysis confirms a possibility of a fundamental increasing of computing performance for this problem by organizing distributed parallel processing. A similar scheme can be used to solve other problems related to handling large volumes of spatial data.

Keywords: 3D reconstruction, hybrid multiprocessing, voxel approach.

Acknowledgements. The work has been partially financially supported by the Programme of the RAS Presidium (project 0262-2014-0157 in the section “Fundamental problems of mathematical modeling”), RFBR (project no. 15-07-00341) and a program “Far East” (project 15-I-4-025).

References

1. Mordohai P., Frahm J.-M., Akbarzadeh A. Real-time video-based reconstruction of urban. *ISPRS Working Group V4 Workshop 3D-ARCH 2007: 3D Virtual Reconstruction and Visualization of Complex Architectures*. ETH Zurich, Switzerland, 2007, pp. 121–131.
2. Curless B., Levoy M. A volumetric method for building complex models from range images. *Conf. Computer Graphics (SIGGRAPH '96)*. 1996, pp. 303–312.
3. Goesele M., Curless B., Seitz S.M. Multi-view stereo revisited. *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*. 2006, vol. 2, pp. 2402–2409.
4. Seitz S., Curless B., Diebel J., Scharstein D., Szeliski R. A comparison and evaluation of multi-view stereo reconstruction algorithms. *IEEE Computer Society Conf. on Vision and Pattern Recognition*. 2006, vol. 1, pp. 519–526.
5. Johnson-Roberson M., Pizarro O., Williams S.B., Mahon I. Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys. *Jour. of Field Robotics, Special Issue: Three-Dimensional Mapping*. Part 3, 2010, vol. 27, iss. 1, pp. 21–51.
6. Bobkov V.A., Kudryashov A.P. Volume approach for the surface reconstruction based on the multiple range-scans. *Informatika i sistemy upravleniya* [Information Science and Control Systems]. 2012, no. 2, pp. 31–38 (in Russ.).
7. Chang J.Y., Park H.S., Park I.K. GPU-friendly multi-view stereo reconstruction using surfel representation and graph cuts. *Computer Vision and Image Understanding*. 2011, vol. 115, iss. 5, pp. 620–634.
8. Mak J., Hess-Flores M., Recker S. GPU-accelerated and efficient multi-view triangulation for scene reconstruction. *IEEE Winter Conf. on Applications of Computer Vision*. Steamboat Springs, CO, 2014, pp. 61–68.
9. Lorensen W.E., Cline H. E., Marching cubes: A high resolution 3D surface construction algorithm. *Conf. Computer Graphics (SIGGRAPH '87)*. 1987, vol. 21, pp. 163–169.
10. Dyken C., Ziegler G. PGU-accelerated data expansion for the Marching Cubes algorithm. *Proc. PGU Technology Conf.* San Jose, CA, 2010, pp. 115–123.

УДК 004.67

DOI: 10.15827/0236-235X.117.081-084

Дата подачи статьи: 24.05.16

2017. Т. 30. № 1. С. 81–84

РЕДАКТИРОВАНИЕ И ВНЕСЕНИЕ ИНФОРМАЦИИ В XML-ДОКУМЕНТЫ АВТОМАТИЗИРОВАННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

А.Н. Трусов, аспирант, TrusovAlexander@hotmail.com;

П.Ю. Иванченко, аспирант, Pavel-Ivanchenko@hotmail.com;

Д.А. Кацуро, аспирант, Davidkacuro@hotmail.com

*(Российский экономический университет им. Г.В. Плеханова (Кемеровский филиал),
Кузнецкий просп., 39, г. Кемерово, 650992, Россия)*

В статье рассматриваются вопросы редактирования и автоматизированного внесения изменений в защищенный от внешнего редактирования конфигурационный файл формата eXtensibleMarkupLanguage (XML) некоторой автоматизированной информационной системы финансово-аналитического содержания. Описываются основная идея и концепция создания модуля для редактирования и внесения информации в исходный XML-файл автоматизированной системы. Рассмотрен способ предоставления конечному пользователю функциональных возможностей информационной системы посредством размещения web-страницы в сети Интернет.

Приведен алгоритм взаимодействия пользователя с модулями программы. Описана подробная техническая реализация алгоритма по редактированию и автоматизированному внесению изменений в конфигурационный файл автоматизированной информационной системы без прямого взаимодействия с самим программным продуктом. Детально проанализирована структура конфигурационного файла и сформированы требования к его формированию. Представлены фрагменты сформированной структуры конфигурационного файла информационной системы, а также программного кода обращения к элементу дерева в XML-файле. Выбрана подходящая программная реализация для внесения социальных и экономических параметров в конфигурационный файл без взаимодействия с программным продуктом.

Потребность в использовании описанного подхода возникает в случаях необходимости оперативной обработки и визуального представления социально-экономической информации на базе ситуационных центров оперативной поддержки принятия экспертных решений в сфере анализа состояния и развития социально-экономических систем. Авторами реализован программный комплекс, состоящий из описываемого модуля и оптимизационной финансово-аналитической автоматизированной информационной системы, который прошел апробацию при решении задач социально-экономического анализа в Ситуационном центре регионального социально-экономического развития Кемеровского филиала Российского экономического университета имени Г.В. Плеханова.

Ключевые слова: XML-документ, web-разработка, ситуационный центр, информационные технологии, автоматизированная информационная система.

В условиях современного информационного общества широкое развитие получили многочисленные системы обработки экономической информации учетного и планирующего характера. Кроме того, появились возможности оперативной обработки и визуального представления информации, что позволяет создавать *ситуационные центры* (СЦ) для поддержки принятия экспертных решений как в online-, так и в offline-режиме. Такая потребность может возникать при использовании программных продуктов в ограниченном функционале, например, при анализе инвестиционных проектов внешним пользователем, а также в условиях СЦ оперативно-экспертной поддержки принятия решений. Одной из важных задач при функционировании СЦ является автоматизированное внесение информации в программные продукты, для которых имеется XML-код [1]. Это может быть обусловлено многими факторами, например, аппаратными требованиями персонального компьютера пользователя, высокой стоимостью полного ПО, необходимостью или возможностью использования только отдельно взятых модулей системы и т.д. Таким образом, возникает необходимость частичного доступа к возможностям некоторой *автоматизированной информационной системы* (АИС)

без ее непосредственной установки на персональный компьютер.

В статье предложены алгоритм, а также техническая реализация процесса удаленного (online) внесения изменений пользователем через определенный требованиями ПО шаблон входной информации, который может располагаться в Excel-документе, web-форме облачного data-центра или в другом электронном источнике.

Основная идея и концепция

Рассмотрим решение сформулированной задачи на примере программной системы «Карма» [2], работа которой описана в [3]. Система представляет собой автоматизированный комплекс для внесения, обработки и анализа входной информации экономического содержания. Данный комплекс имеет возможности создания и корректировки математических моделей в форме многопараметрических задач линейного программирования, контроля корректности внесения информации, создания собственной конфигурации проекта (разделов, блоков переменных и т.п.), а также графический анализатор, визуализирующий многопараметрические зависимости и Парето-множества. Это

позволяет использовать данную систему специалисту-математику, экономисту-аналитику и бизнесмену [4–9]. Однако для пользователя эта система представляется в виде XML-файла, непосредственное изменение которого ему недоступно, и не позволяет автоматизированно вносить массивы входной информации для использования функциональных возможностей системы.

Рассмотрим следующий вариант получения входной информации от пользователя в указанный пакет. Информация об экономических характеристиках некоторого инвестиционного проекта заносится в файл строго фиксированного формата (Word, Excel или др.) и передается в пакет «Карма» путем ее преобразования через разработанный авторами специальный программный модуль – Instrument for Data Acquisition (IDA-модуль), работа которого будет описана далее. На рисунке представлена схема преобразования информации проектов социально-экономических систем от стадии ее внесения в исходный XML-файл до выдачи результатов расчетов.

Рассмотрим подробно каждый из этапов.

- Для работы с IDA пользователю необходимо внести в соответствующие поля стандартной формы IDA входную информацию о характеристиках инвестиционного, производственного или финансового проекта.
- IDA-модуль автоматически вносит отредактированные значения в XML-файл, осуществляя подстановку введенных параметров в соответствующую группу показателей, необходимых для работы системы. При этом последовательность внесения информации через IDA соответствует последовательности ее обработки в XML-файле. То есть по завершении процесса внесения данных в IDA вся информация попадает в облачное хранилище, откуда с помощью административной панели IDA вносится в необходимый XML-файл для дальнейших расчетов оператором.
- Результаты многопараметрического анализа проекта предоставляются заказчику в виде отчета.

Техническая реализация

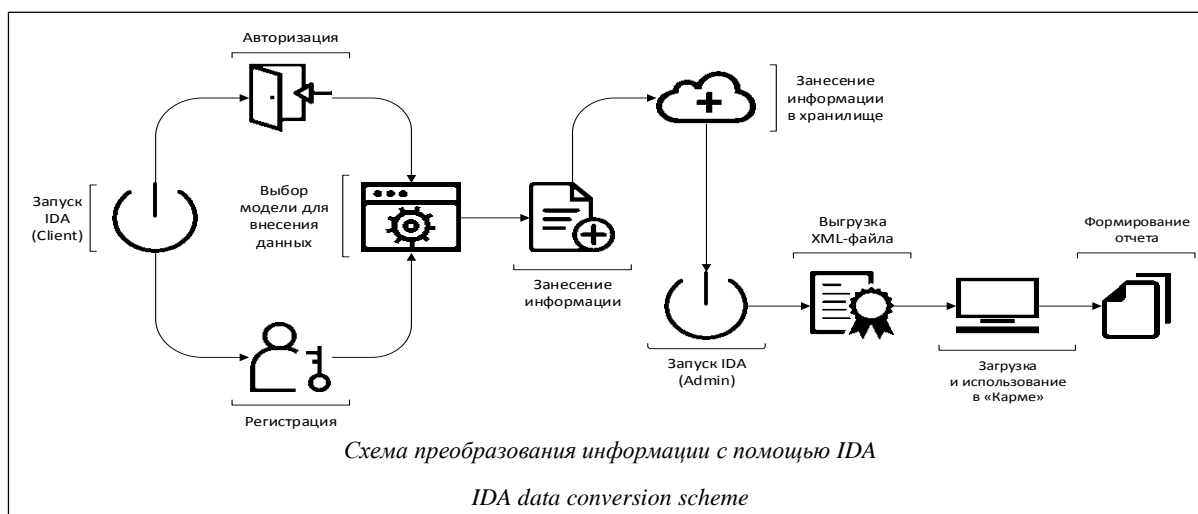
В основе загрузки параметров для расчета моделей в [2] лежит хранение параметров в виде XML-файлов. Конфигурационный файл содержит экономические параметры, разделенные на соответствующие группы по принадлежности (например, «Налоги», «ФОТ и другие затраты», «ИП внешние характеристики» и т.д.). Он защищен от редактирования с помощью внутренних инструментов. При занесении информации следует учитывать повышенный риск возникновения ошибок некорректного ввода значений в XML-файл.

В процессе изучения конфигурационного файла первоначально необходимо обратить внимание на его структуру, включающую элементы разметки (markup), содержимое файла, а также XML-тэги, предназначенные для определения элементов документа, их атрибутов и других конструкций языка.

XML-файл имеет древовидную структуру. В документе всегда есть корневой элемент. У элемента дерева всегда существуют потомки и предки, кроме корневого элемента, у которого предков нет, и тупиковых элементов (листьев дерева), у которых нет потомков [10, 11].

В процессе изучения вопроса были сформулированы следующие требования к формированию XML-файла:

- каждый открывающий XML-тэг определяет некоторую область данных в документе и должен иметь своего закрывающего «напарника»;
- в XML-файле учитывается регистр символов;
- все значения атрибутов, используемых в определении тэгов, должны быть заключены в кавычки;
- вложенность тэгов в XML-файле строго контролируется, поэтому необходимо следить за очередностью открывающих и закрывающих тэгов;
- вся информация, располагающаяся между начальным и конечными тэгами, рассматривается в XML-файле как данные, и поэтому учитываются



все символы форматирования (пробелы, переводы строк, табуляции не игнорируются, как в HTML).

После соблюдения этих правил документ принято считать формально правильным и все анализаторы, предназначенные для разбора XML-документов, смогут работать корректно [10, 11]. Однако очень важно отметить, что, помимо проверки на формальное соответствие грамматике языка, в XML-файле могут присутствовать средства контроля над его содержанием, за соблюдением правил, определяющих необходимые соотношения между элементами и формирующими структуру XML-файла. Чтобы обеспечить проверку корректности XML-файла, необходимо использовать анализаторы, производящие подобную проверку. На сегодняшний день практикуются два способа контроля соответствия XML-файла вышеописанным правилам: DTD-определения (DocumentTypeDefinition) и схемы данных (SemanticSchema) [10, 11]. Представим фрагмент конфигурационного файла информационной системы «Карма» после завершения вышеописанных действий:

```
<category groupname="ИП внешние характеристики" group
title="ИП внешние характеристики" expert_mode_only="0">
  <parameters>
    <parameter type="0" param_name="N" param_title=
"количество видов продукции">3</parameter>
    <parameter type="0" param_name="N1" param_title=
"количество направлений риска">2</parameter>
```

По завершении анализа конфигурационного файла необходимо выбрать подходящую программную реализацию для внесения экономических параметров в конфигурационный файл без взаимодействия с «Кармой».

Наиболее подходящим программным методом решения данной проблемы явился язык запросов к элементам XML-файла – XPath (XML Path Language). Он был разработан для организации доступа к частям документа формата XML в файлах трансформации XSLT (eXtensibleStylesheetLanguage-Transformations) и является стандартом консорциума W3C [10–12].

На каждом шаге отбираются элементы дерева XML-файла, соответствующие последовательности обращения к параметрам модели «Карма». В результате формируется множество элементов дерева, отвечающих структуре входной информации [2], которые могут использоваться для проведения расчетов в нем. Представим фрагмент программного кода «обращение к элементу дерева в XML-файле»:

```
protected void xmlTreeView_SelectedNodeChanged(object
sender, EventArgs e)
{
  XmlDocument doc = new XmlDocument();
  doc.Load(Server.MapPath(Session["CurrentXMLFile"]].
ToString());
  XmlNode node = doc.SelectSingleNode(xmlTreeView.
SelectedNode.DataPath);
  txtEditXML.Text = node.InnerXml;
  Session["CurrentEditXPath"] = xmlTreeView.SelectedNode.
DataPath;
}
```

Заключение

В результате проведенной работы удалось реализовать программный модуль для внесения параметров в конфигурационный файл автоматизированной информационной системы финансово-аналитического содержания [2]. В перспективе данное решение предоставляет возможность конечному пользователю взаимодействовать с системой «Карма» без непосредственной установки ее на персональном компьютере, обеспечивая при этом доступ к функциональным возможностям автоматизированной информационной системы для проведения инвестиционного, производственного и финансового анализа проектов развития социально-экономических систем.

Отметим, что предложенный комплекс, состоящий из финансово-аналитической системы и пакета автоматического внесения информации в нее, ориентирован на использование в СЦ социально-экономического анализа, является удобным инструментом оперативной поддержки экспертных решений в случае как очного, так и удаленного присутствия экспертов. Данный комплекс прошел тестирование в СЦ регионального социально-экономического развития Кемеровского филиала Российского экономического университета имени Г.В. Плеханова.

Литература

1. Белорусов А. И. Интеграция информационных систем на основе стандартов XML и WEB-сервисов в сфере закупок // Молодой ученый. 2015. № 11. С. 9–15.
2. Медведев А.В., Победаш П.Н., Смольянинов А.В., Горбунов М.А. Конструктор и решатель дискретных задач оптимального управления («Карма»). Свид. о регистр. прог. для ЭВМ в Роспатенте № 2008614387 от 11.09.2008.
3. Горбунов М.А., Медведев А.В., Победаш П.Н., Смольянинов А.В. Оптимизационный пакет прикладных программ «Карма» и его применение в задачах бизнес-планирования // Фундаментальные исследования. 2015. № 4. С. 42–47.
4. Трусев А.Н., Иванченко П.Ю., Кацуро Д.А. Разработка автоматизированной информационной системы для оптимизационного анализа экономических процессов // Современные наукоемкие технологии. 2014. № 11. С. 38–40.
5. Медведев А.В., Трусев А.Н. Алгоритм автоматизированного внесения региональной экономико-статистической информации в XML-файлы аналитической информационной системы на основе Excel-надстроек // Междунар. журн. прикладных и фундаментальных исследований. 2016. № 3 (Ч. 4). С. 525–528.
6. Трусев А.Н. Использование автоматизированных информационных систем для решения оптимизационных задач инвестиционного планирования // Новая наука: теоретический и практический взгляд: матер. Междунар. науч.-практич. конф. Стерлитамак: Изд-во АМИ, 2016. Ч. 2. С. 145–147.
7. Медведев А.В. Поддержка принятия решений при управлении экономикой региона: монография. Кемерово: Изд-во КемГУ, 2011. 106 с.
8. Медведев А.В. Оптимизационная система поддержки принятия решений в бизнес-планировании // Успехи современного естествознания. 2015. № 1 (4). С. 679–683.
9. Медведев А.В. Об эффективном инструментарии анализа экономических систем // Междунар. журн. прикладных и фундаментальных исследований. 2014. № 11 (5). С. 763–766.
10. The main international standards organization. World Wide Web Consortium (W3C): XMLPathLanguage (XPath) 2.0. URL:

<http://www.w3.org/TR/xpath20/> (дата обращения: 01.04.2016).

11. XPath Tutorial. URL: http://www.w3schools.com/xml/xpath_intro.asp (дата обращения: 29.03.2016).

12. Березкин Д.В. Метод автоматизированного извлечения знаний из слабоструктурированных источников и его применение для создания корпоративных информационных систем.

URL: <http://www.raai.org/resurs/papers/kolonna2009/> (дата обращения: 11.04.2016).

13. Трусов А.Н., Кацуро Д.А., Иванченко П.Ю. Извлечение и обработка информации с сайтов экономической статистики // Междунар. журн. приклад. и фундамент. исследований. 2013. № 12. С. 120–122.

Software & Systems

DOI: 10.15827/0236-235X.117.081-084

Received 24.05.16

2017, vol. 30, no. 1, pp. 81–84

EDITING AND ENTERING INFORMATION INTO XML-DOCUMENTS OF AUTOMATED INFORMATION SYSTEMS

A.N. Trusov¹, Postgraduate Student, TrusovAlexander@hotmail.com

P.Yu. Ivanchenko¹, Postgraduate Student, Pavel-Ivanchenko@hotmail.com

D.A. Katsuro¹, Postgraduate Student, Davidkacuro@hotmail.com

¹ Plekhanov Russian University of Economics, Kemerovo Institute (branch), Kuznetsky Ave. 39, Kemerovo, 650992, Russian Federation

Abstract. The article considers the issues of automated editing and amending changes to a eXtensibleMarkupLanguage format (XML) configuration file protected from external editing in some automated information system (AIS) with financial and analytical content. It describes the basic idea and the concept of creating a module for editing and entering the information in an original XML-file in an automated system. It considers a method of providing information system functionality to an end-user by placing a web page on the Internet.

The paper shows the algorithm of interaction between a user and program modules. It also describes in detail the technical implementation of the algorithm for editing and automated changing an AIS configuration file without direct interaction with software. The article analyzes the configuration file structure in detail and forms the requirements for its creation. It presents fragments of the configuration file formed structure in the information system, as well as the code referring to the tree element in the XML-file. The authors select the appropriate software implementation to enter social and economic parameters into a configuration file without interaction with a software product.

The described approach is necessary when there is a need in operational processing and visual representation of socio-economic information based on situational centers of operational support decision-making in expert analysis of the state and development of socio-economic systems. The authors implemented a software package consisting of the described module and the optimization of financial and analytical AIS, which has been tested in solving problems of social and economic analysis of the situation center in the regional socio-economic development of the Kemerovo branch of the Russian University of Economics named after G.V. Plekhanov.

Keywords: XML-document, Web-development, situation center, information technology, automated information system (AIS).

References

1. Belorусov A.I. Integration of information systems based on XML standards and WEB-services in the case of municipal procurement. *Molodoy ucheny*, 2015, no. 11, pp. 9–15.
2. Medvedev A.V., Pobedash P.N., Smolyaninov A.V., Gorbunov M.A. *Konstruktor i reshatel diskretnykh zadach optimalnogo upravleniya ("Karma")*. [A Constructor and Solver for Discrete Optimal Control Problems ("Karma")]. Software State Registration Certificate no. 2008614387. 2008.
3. Gorbunov M.A., Medvedev A.V., Pobedash P.N., Smolyaninov A.V. Optimization software package "Karma" and its application in business planning problems. *Fundamentalnye issledovaniya* [Fundamental research]. 2015, no. 4, pp. 42–47 (in Russ.).
4. Trusov A.N., Ivanchenko P.Yu., Katsuro D.A. Development of an automated information system for the optimization analysis of economic processes. *Sovremennye naukoemkie tekhnologii* [Modern High Technologies]. 2014, no. 11, pp. 38–40 (in Russ.).
5. Medvedev A.V., Trusov A.N. An algorithm for automated application of regional economic and statistical data into XML-files of analytical information system based on Excel add-ons. *Mezhdunar. zhurn. prikladnykh i fundamentalnykh issledovaniy* [Int. Jour. of Applied and Basic Research]. 2016, no. 3 (p. 4), pp. 525–528 (in Russ.).
6. Trusov A.N. Using automated information systems to solve optimization problems of investment planning. *Novaya nauka: teoretichesky i praktichesky vzglyad: mater. Mezhdunar. nauch.-praktich. konf.* [Proc. Int. Science and Practice Conf. "New Science: Theoretical and Practical View"]. Sterlitamak, AMI Publ., 2016, part 2, pp. 145–147 (in Russ.).
7. Medvedev A.V. *Podderzhka prinyatiya resheny pri upravlenii ekonomikoy regiona* [Decision-Making Support when Managing Regional Economy]. Monograph. Kemerovo, KemSU Publ., 2011, 106 p.
8. Medvedev A.V. An optimization decision support system in business planning. *Uspekhi sovremennogo estestvoznaniya* [Advances in Current Natural Sciences]. 2015, no. 1 (4), pp. 679–683 (in Russ.).
9. Medvedev A.V. On effective toolkit of economic system analysis. *Mezhdunar. zhurn. prikladnykh i fundamentalnykh issledovaniy* [Int. Jour. of Applied and Basic Research]. 2014, no. 11 (5), pp. 763–766 (in Russ.).
10. The Main International Standards Organization. World Wide Web Consortium (W3C): *XMLPathLanguage (XPath) 2.0*. Available at: <http://www.w3.org/TR/xpath20/> [accessed April 1, 2016].
11. XPath Tutorial. Available at: http://www.w3schools.com/xml/xpath_intro.asp [accessed March 29, 2016].
12. Berезкин D.V. *Method for automated knowledge extraction from semi-structured sources and its application to corporate information systems*. Available at: <http://www.raai.org/resurs/papers/kolonna2009/> [accessed April 11, 2016].
13. Trusov A.N., Katsuro D.A., Ivanchenko P.Yu. Extracting and processing data from sites of economic statistics. *Mezhdunar. zhurn. prikladnykh i fundamentalnykh issledovaniy* [Int. Journ. of Applied and Basic Research]. 2013, no. 12, pp. 120–122 (in Russ.).

УДК 004.048

DOI: 10.15827/0236-235X.117.085-099

Дата подачи статьи: 19.07.16

2017. Т. 30. № 1. С. 85–99

МЕТОДЫ АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ ТЕКСТОВ

*Т.В. Батура, к.ф.-м.н., ведущий научный сотрудник, tatiana.v.batura@gmail.com
(Новосибирский государственный университет, ул. Пирогова, 2, г. Новосибирск, 630090, Россия);
старший научный сотрудник (Институт систем информатики им. А.П. Еришова СО РАН,
просп. Лаврентьева, 6, г. Новосибирск, 630090, Россия)*

Классификация текстов является одной из основных задач компьютерной лингвистики, поскольку к ней сводится ряд других задач: определение тематической принадлежности текстов, автора текста, эмоциональной окраски высказываний и др. Для обеспечения информационной и общественной безопасности большое значение имеет анализ в телекоммуникационных сетях контента, содержащего противоправную информацию (в том числе данные, связанные с терроризмом, наркоторговлей, подготовкой протестных движений или массовых беспорядков).

Данная статья представляет собой обзор методов классификации текстов, целями которого являются сравнение современных методов решения задачи классификации текстов, обнаружение тенденций развития данного направления, а также выбор наилучших алгоритмов для применения в исследовательских и коммерческих задачах.

Широко известный современный подход к классификации основывается на методах машинного обучения. В данной статье описываются наиболее распространенные алгоритмы построения классификаторов, проводимые с ними эксперименты и результаты этих экспериментов. Обзор подготовлен на основе выполненных за 2011–2016 гг. научных работ, находящихся в открытом доступе в сети Интернет и опубликованных в авторитетных журналах или в трудах международных конференций, высоко оцениваемых научным сообществом.

В статье произведены анализ и сравнение качества работы различных методов классификации по таким характеристикам, как точность, полнота, время работы алгоритма, возможность работы алгоритма в инкрементном режиме, количество предварительной информации, необходимой для классификации, независимость от языка.

Ключевые слова: классификация текстов, анализ текстовой информации, обработка данных, машинное обучение, нейронные сети, качество классификации.

Прогресс в области микроэлектроники и информационных технологий обусловил широкое распространение обработки в реальном времени больших потоков данных. Например, многие простые операции повседневной жизни, такие как использование кредитной карты или телефона, требуют автоматизированного создания, анализа и обработки различных данных. Поскольку эти операции часто выполняются большим числом участников, необходимы распределенные и массовые потоки данных. Точно так же социальные сети содержат большое количество специфических сетевых и текстовых потоков данных. Поэтому актуальна проблема создания моделей и алгоритмов, позволяющих эффективно обрабатывать большие потоки данных, особенно в условиях ограниченных временных и других ресурсов.

Для обеспечения информационной и общественной безопасности важное значение имеет анализ в телекоммуникационных сетях контента, содержащего противоправную информацию (в том числе данные, связанные с терроризмом, наркоторговлей, сетевым экстремизмом, подготовкой протестных движений или массовых беспорядков).

Целями данного обзора являются сравнение современных методов решения задачи классификации текстов, обнаружение тенденций развития данного направления, а также выбор наилучших алгоритмов для применения в исследовательских и коммерческих задачах.

Методы классификации текстов лежат на стыке двух областей – информационного поиска и машинного обучения. Их сходство состоит в способах

представления самих документов и способах оценки качества алгоритмов. На сегодняшний день разработано большое количество методов и их различных вариаций для классификации текстов. Каждая группа методов имеет свои преимущества и недостатки, области применения, особенности и ограничения.

Особый интерес представляет случай, когда данные поступают в виде потока, например в телекоммуникационных сетях. Определенные трудности возникают из-за того, что обучение модели всегда основывается на совокупности свойств набора документов. Эти совокупные свойства могут изменяться с течением времени, и при построении потокового классификатора необходимо учитывать возможные изменения исходного распределения данных [1]. Желательно, чтобы выбранный метод мог поддерживать инкрементное обучение, то есть чтобы классификатор обучался на каждом отдельно взятом образце в режиме реального времени. При инкрементном обучении обучающие примеры поступают последовательно в процессе работы алгоритма, так что классификатор должен постоянно корректировать результаты обучения и дообучаться. При неинкрементном обучении вся обучающая выборка предоставляется сразу полностью. Ясно, что в случае инкрементного обучения поведение классификатора в процессе работы меняется, что уменьшает его предсказуемость и может осложнить настройку системы. В то же время инкрементное обучение делает систему гораздо более гибкой, адаптируемой к изменяющимся условиям.

Особенности процесса классификации в потоке связаны еще с тем, что не всегда удастся контролировать скорость поступления данных. Некоторые классы документов могут встречаться в потоке только время от времени. Обнаружить этот редкий класс бывает непросто, и классификация текстов в таких случаях становится чрезвычайно сложной задачей.

Сравнение методов построения классификаторов является довольно сложной задачей по причине того, что разные входные данные могут приводить к различным результатам. Поэтому необходимо осуществить их программную реализацию и вычисление эффективности на одинаковых наборах документов для обучения и тестирования.

Формальная постановка задачи классификации текстов

Следует отличать классификацию от кластеризации. При классификации документов категории определены заранее, при кластеризации они не заданы и даже информация об их количестве может отсутствовать.

Формально постановку задачи классификации можно записать следующим образом.

Имеются множество документов $D = \{d_1, \dots, d_{|D|}\}$ и множество возможных категорий (классов) $C = \{c_1, \dots, c_{|C|}\}$. Неизвестная целевая функция $\Phi: D \times C \rightarrow \{0, 1\}$ задается формулой

$$\Phi(d_j, c_i) = \begin{cases} 0, & \text{если } d_j \notin c_i, \\ 1, & \text{если } d_j \in c_i. \end{cases} \quad (1)$$

Требуется построить классификатор Φ' , максимально близкий к Φ .

В такой постановке задачи следует отметить, что о категориях и документах нет никакой дополнительной информации, кроме той, которую можно извлечь из самого документа.

Если классификатор выдает точный ответ:

$$\Phi': D \times C \rightarrow \{0, 1\}, \quad (2)$$

то классификация называется точной.

Если классификатор определяет степень подобия (Categorization Status Value) документа:

$$CSV: D \rightarrow [0, 1], \quad (3)$$

то классификация называется пороговой.

В общем случае процесс обучения с учителем (обучение по прецедентам, supervised learning) заключается в следующем. Системе предъявляется набор примеров, связанных с какой-либо заранее неизвестной закономерностью. Этот набор иногда называют обучающей выборкой L . Ее используют для обучения классификатора и определения значения его параметров, при которых классификатор выдает лучший результат. Далее в системе вырабатываются решающие правила, с помощью которых происходит разделение множества примеров на заданные классы. Качество разделения проверяется

тестовой выборкой примеров T . При этом необходимо, чтобы выполнялись условия

$$L \cap T = \emptyset, \quad (4)$$

$$\Omega = L \cup T \subset C \times D. \quad (5)$$

Для множества примеров Ω известны значения целевой функции Φ .

Если в задаче каждому документу $d \in D$ может соответствовать только одна категория $c \in C$, то имеет место однозначная классификация, а если произвольное количество категорий, то многозначная классификация.

Частным случаем однозначной классификации является бинарная классификация, когда коллекцию документов нужно разбить на две непересекающиеся категории. Например, задача определения тональности высказываний (положительная или отрицательная окраска) или задача обнаружения спама (является сообщение спамом или нет) решается при помощи бинарного классификатора.

Решение задачи классификации состоит из четырех последовательных этапов:

- предобработка и индексация документов;
- уменьшение размерности пространства признаков;
- построение и обучение классификатора с помощью методов машинного обучения;
- оценка качества классификации.

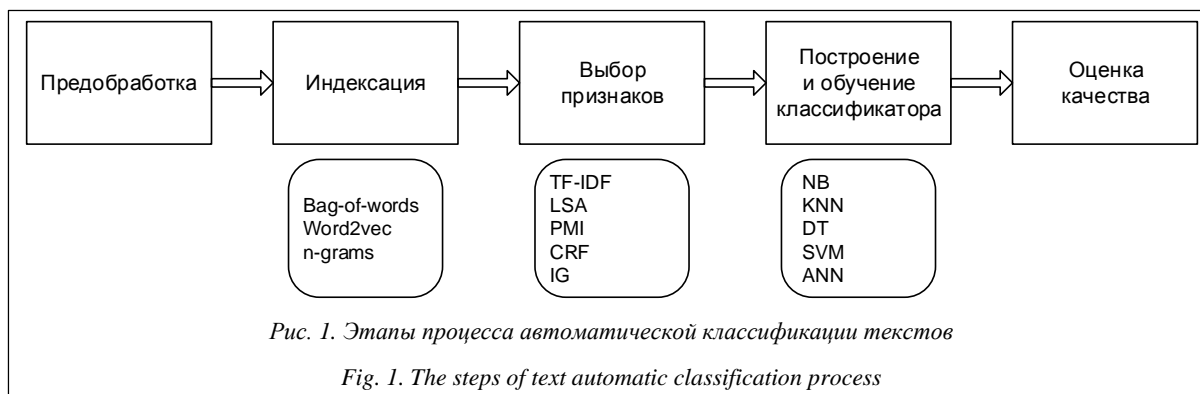
При выборе конкретного алгоритма классификации следует учитывать особенности каждого из них. По-прежнему остается нерешенным вопрос определения набора классифицирующих признаков, их количества и способов вычисления весов. В алгоритмах глубокого обучения точность классификации сильно зависит от наличия обучающей выборки подходящего размера. Подготовка такой выборки – очень трудоемкий процесс. До сих пор остается также открытой проблема подбора параметров некоторых алгоритмов на этапе обучения.

Далее подробно рассмотрен каждый из этапов, описаны различные алгоритмы построения классификаторов, проводимые с ними эксперименты и результаты этих экспериментов.

Описание методов классификации

На рисунке 1 представлена общая схема процесса классификации. Рассмотрим каждый из его этапов.

Предобработка и индексация документов. Предварительная обработка текста включает в себя токенизацию, удаление функциональных слов (семантически нейтральных слов, таких как союзы, предлоги, артикли и пр.). Далее осуществляется морфологический анализ (производятся разметка по частям речи и стемматизация). Это позволяет значительно сократить размерность пространства. В результате в качестве признаков документа выступают все значимые слова, встречающиеся в документе.



Индексация документов – это построение некоторой числовой модели текста, которая переводит текст в удобное для дальнейшей обработки представление.

Например, модель «мешка слов» (bag-of-words) позволяет представить документ в виде многомерного вектора слов и их весов в документе [2]. Другими словами, каждый документ – это вектор в многомерном пространстве, координаты которого соответствуют номерам слов, а значения координат – значениям весов.

Другая распространенная модель индексации – Word2vec [3]. Она представляет каждое слово в виде вектора, который содержит информацию о контекстных (сопутствующих) словах.

Еще одна модель индексации основана на учете n -грамм [2], то есть последовательностей из соседних символов.

Очевидно, что для обучающих и тестовых документов должен применяться один и тот же метод индексации.

Уменьшение размерности пространства признаков. Вычислительная сложность различных методов классификации напрямую зависит от размерности пространства признаков. Поэтому для эффективной работы классификатора часто прибегают к сокращению числа используемых признаков (терминов).

За счет уменьшения размерности пространства терминов можно снизить эффект переобучения – явление, при котором классификатор ориентируется на случайные или ошибочные характеристики обучающих данных, а не на важные и значимые. Переобученный классификатор хорошо работает на тех экземплярах, на которых он обучался, и значительно хуже на тестовых данных. Чтобы избежать переобучения, количество обучающих примеров должно быть соразмерно числу используемых терминов. В некоторых случаях сокращение размерности пространства признаков в 10 раз (и даже в 100) может приводить лишь к незначительному ухудшению работы классификатора.

Существуют несколько способов определения веса признаков документа. Наиболее распространенный – вычисление функции TF-IDF [2, 4, 5]. Его основная идея состоит в том, чтобы больший вес

получали слова с высокой частотой в пределах конкретного документа и с низкой частотой употреблений в других документах.

Вычисляется частота термина TF (term frequency) – оценка важности слова в пределах одного документа d по формуле

$$TF = n_{t,d} / n_d, \quad (6)$$

где $n_{t,d}$ – количество употреблений слова t в документе d ; n_d – общее число слов в документе d .

Обратная частота документа IDF (inverse document frequency) – инверсия частоты, с которой слово встречается в документах коллекции. IDF уменьшает вес общеупотребительных слов по формуле

$$IDF = \log(|D| / D_t), \quad (7)$$

где $|D|$ – общее количество документов в коллекции; D_t – количество всех документов, в которых встречается слово t .

Итоговый вес термина в документе относительно всей коллекции документов вычисляется по формуле

$$V_{t,d} = TF \cdot IDF. \quad (8)$$

Следует отметить, что по формуле (8) оценивается значимость термина только с точки зрения частоты вхождения в документ, без учета порядка следования терминов в документе и их лексической сочетаемости.

Для уменьшения размерности пространства терминов также применяют латентно-семантический анализ (LSA), использующий сингулярное разложение матриц [3, 6], поточечную взаимную информацию (PMI) [6, 7] (разновидность ассоциативной меры), условные случайные поля (CRF) [8] (обобщение скрытой марковской модели). Встречаются исследования [4, 9], в которых применяются статистические критерии и относительная энтропия для вероятностных распределений, называемая коэффициентом усиления информации, или дивергенцией Кульбака–Лейблера.

Построение и обучение классификатора с помощью методов машинного обучения. Можно выделить следующие методы классификации:

- вероятностные (например NB [4, 6]);
- метрические (например KNN [9]);
- логические (например DT [6, 10]);

- линейные (например SVM [4, 5, 6, 9]; логистическая регрессия [2, 8, 10]);
- методы на основе искусственных нейронных сетей (например FFBP [4, 10], RNN [8], DAN2 [9], CNN [2]).

Далее обобщенно описываются эти методы, указываются преимущества и недостатки каждого из них.

Метод Байеса (Naive Bayes, NB) относится к вероятностным методам классификации.

Пусть $P(c_i/d)$ – вероятность того, что документ, представленный вектором $d = (t_1, \dots, t_n)$, соответствует категории c_i для $i = 1, \dots, |C|$. Задача классификатора заключается в том, чтобы подобрать такие значения c_i и d , при которых значение вероятности $P(c_i/d)$ будет максимальным:

$$CSV(d) = \arg \max_{c_i \in C} P(c_i | d). \quad (9)$$

Для вычисления значений $P(c_i/d)$ пользуются теоремой Байеса:

$$P(c_i | d) = \frac{P(c_i)P(d | c_i)}{P(d)}, \quad (10)$$

где $P(c_i)$ – априорная вероятность того, что документ отнесен к категории c_i ; $P(d | c_i)$ – вероятность найти документ, представленный вектором $d = (t_1, \dots, t_n)$, в категории c_i ; $P(d)$ – вероятность того, что произвольно взятый документ можно представить в виде вектора признаков $d = (t_1, \dots, t_n)$.

По сути $P(c_i)$ является отношением количества документов из обучающей выборки L , отнесенных в категорию c_i , к количеству всех документов из L .

$P(d)$ не зависит от категории c_i , а значения t_1, \dots, t_n заданы заранее, поэтому знаменатель – это константа, не влияющая на выбор наибольшего из значений $P(c_i/d)$.

Вычисление $P(d | c_i)$ затруднительно из-за большого количества признаков t_1, \dots, t_n , поэтому делают «наивное» предположение о том, что любые две координаты, рассматриваемые как случайные величины, статистически не зависят друг от друга. Тогда можно воспользоваться формулой

$$P(d | c_i) = \prod_{k=1}^n P(t_k | c_i). \quad (11)$$

Далее все вероятности подсчитываются по методу максимального правдоподобия.

Преимущества метода:

- высокая скорость работы;
- поддержка инкрементного обучения;
- относительно простая программная реализация алгоритма;
- легкая интерпретируемость результатов работы алгоритма.

Недостатки метода: относительно низкое качество классификации и неспособность учитывать зависимость результата классификации от сочетания признаков.

Метод k ближайших соседей (k Nearest Neighbors, KNN) относится к метрическим мето-

дам классификации. Чтобы найти категорию, соответствующую документу d , классификатор сравнивает d со всеми документами из обучающей выборки L , то есть для каждого $d_z \in L$ вычисляется расстояние $\rho(d_z, d)$. Далее из обучающей выборки выбираются k документов, ближайших к d . Согласно методу k ближайших соседей, документ d считается принадлежащим тому классу, который является наиболее распространенным среди соседей данного документа, то есть для каждого класса c_i вычисляется функция ранжирования:

$$CSV(d) = \sum_{d_z \in L_k(d)} \rho(d_z, d) \cdot \Phi(d_z, c_i), \quad (12)$$

где $L_k(d)$ – ближайшие k документов из L к d ; $\Phi(d_z, c_i)$ – известные величины, уже расклассифицированные по категориям документы обучающей выборки.

Преимущества метода:

- возможность обновления обучающей выборки без переобучения классификатора;
- устойчивость алгоритма к аномальным выбросам в исходных данных;
- относительно простая программная реализация алгоритма;
- легкая интерпретируемость результатов работы алгоритма;
- хорошее обучение в случае с линейно неразделимыми выборками.

Недостатки метода:

- репрезентативность набора данных, используемого для алгоритма;
- высокая зависимость результатов классификации от выбранной метрики;
- большая длительность работы из-за необходимости полного перебора обучающей выборки;
- невозможность решения задач большой размерности по количеству классов и документов.

Метод деревьев решений (Decision Trees, DT) относится к логическим методам классификации.

Деревом решений называют ациклический граф, по которому производится классификация объектов (в нашем случае текстовых документов), описанных набором признаков. Каждый узел дерева содержит условие ветвления по одному из признаков. У каждого узла столько ветвлений, сколько значений имеет выбранный признак. В процессе классификации осуществляются последовательные переходы от одного узла к другому в соответствии со значениями признаков объекта. Классификация считается завершённой, когда достигнут один из листьев (конечных узлов) дерева. Значение этого листа определит класс, которому принадлежит рассматриваемый объект. На практике обычно используют бинарные деревья решений, в которых принятие решения перехода по ребрам осуществляется простой проверкой наличия признака в документе. Если значение признака

меньше определенного значения, выбирается одна ветвь, если больше или равно, другая.

В отличие от остальных подходов, представленных ранее, подход, использующий деревья решений, относится к символьным (то есть нечисловым) алгоритмам.

Алгоритм построения бинарного дерева решений состоит из следующих шагов.

Создается первый узел дерева, в который входят все документы, представленные всеми имеющимися признаками. Размер вектора признаков для каждого документа равен n , так как $d = (t_1, \dots, t_n)$.

Для текущего узла дерева выбираются наиболее подходящий признак t_k и его наилучшее пограничное значение v_k .

На основе пограничного значения выбранного признака производится разделение обучающей выборки на две части. Далее выбранный признак не включается в описание фрагментов в этих частях, то есть фрагменты в частях представляются вектором с размерностью $n - 1$.

Образовавшиеся подмножества обрабатываются аналогично до тех пор, пока в каждом из них не останутся документы только одного класса или признаки для различения документов.

Когда говорят о выборе наиболее подходящего признака, как правило, подразумевают частотный признак, то есть любой признак текста, допускающий возможность нахождения частоты его появления в тексте. Лучшим для разделения является признак, дающий максимальную на данном шаге информацию о категориях. Таким признаком для текста может являться, например, ключевое слово. С этой точки зрения любой частотный признак можно считать переменной. Тогда выбор между двумя наиболее подходящими признаками сводится к оценке степени связанности двух переменных. Поэтому для выбора подходящего признака на практике применяют различные критерии проверки гипотез, то есть критерии количественной оценки степени связанности двух переменных, поставленных во взаимное соответствие, где 0 соответствует полной независимости переменных, а 1 – их максимальной зависимости.

Для исследования связи между двумя переменными удобно использовать представление совместного распределения этих переменных в виде таблицы сопряженности (факторной таблицы, или матрицы частот появления признаков). Она является наиболее универсальным средством изучения статистических связей, так как в ней могут быть представлены переменные с любым уровнем измерения. Таблицы сопряженности часто используются для проверки гипотезы о наличии связи между двумя признаками при помощи различных статистических критериев: критерия Фишера (точного теста Фишера), критерия согласия Пирсона (критерия хи-квадрат), критерия Крамера, критерия Стьюдента (t-критерия Стьюдента) и пр.

Преимущества метода:

- относительно простая программная реализация алгоритма;
- легкая интерпретируемость результатов работы алгоритма.

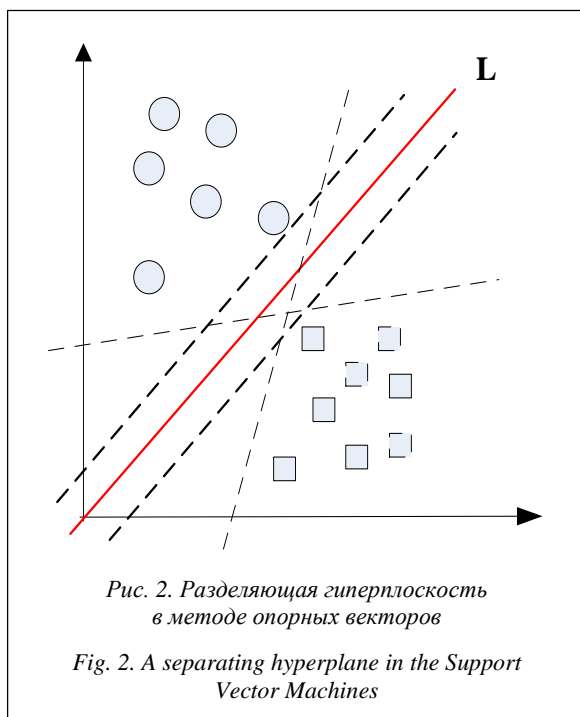
Недостатки метода: неустойчивость алгоритма по отношению к выбросам в исходных данных и большой объем данных для получения точных результатов.

Метод опорных векторов (Support Vector Machine, SVM) является линейным методом классификации. В настоящее время этот метод считается одним из лучших. Рассмотрим множество документов, которые необходимо расклассифицировать. Сопоставим ему множество точек в пространстве размерности $|D|$.

Выборку точек называют линейно разделимой, если принадлежащие разным классам точки можно разделить с помощью гиперплоскости (в двухмерном случае гиперплоскостью является прямая линия). Очевидный способ решения задачи в таком случае – провести прямую так, чтобы по одну сторону от нее лежали все точки одного класса, а по другую – все точки другого класса. Тогда для классификации неизвестных точек достаточно будет посмотреть, с какой стороны прямой они окажутся.

В общем случае можно провести бесконечное множество гиперплоскостей (прямых), удовлетворяющих нашему условию. Ясно, что лучше всего выбрать прямую, максимально удаленную от имеющихся точек. В методе опорных векторов расстоянием между прямой и множеством точек считается расстояние между прямой и ближайшей к ней точкой из множества. Именно такое расстояние и максимизируется в данном методе. Гиперплоскость, максимизирующая расстояние до двух параллельных гиперплоскостей, называется разделяющей (на рисунке 2 обозначена буквой L). Ближайшие к параллельным гиперплоскостям точки называются опорными векторами (рис. 2), через них проходят пунктирные линии. Другими словами, алгоритм работает в предположении, что, чем больше разница или расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора, так как максимизация зазора между классами способствует более уверенной классификации.

На практике структура данных зачастую бывает неизвестна и очень редко удается построить разделяющую гиперплоскость, а значит, невозможно гарантировать линейную разделимость выборки. Могут существовать такие документы, которые алгоритм отнесет к одному классу, а в действительности они должны относиться к противоположному. Такие данные называются выбросами, они создают погрешность метода, поэтому было бы лучше их игнорировать. В этом заключается суть проблемы линейной неразделимости.



Выборку называют линейно неразделимой, если точки, принадлежащие разным классам, нельзя разделить с помощью гиперплоскости. Когда такой разделяющей гиперплоскости не существует, необходимо перейти от исходного пространства признаков документов к новому, в котором обучающая выборка окажется линейно разделимой. Для этого каждое скалярное произведение необходимо заменить на некоторую функцию, отвечающую определенным требованиям. Например, можно назначать некий штраф за каждый неверно расклассифицированный документ. Эту функцию называют ядром. Замена скалярного произведения функцией-ядром позволяет перейти к другому пространству признаков, где данные уже будут разделимы.

В случае линейной неразделимости проблема поиска оптимальной разделяющей гиперплоскости сводится к задаче, эквивалентной поиску седловой точки функции Лагранжа с условиями дополняющей нежесткости. Полученная система уравнений решается методами квадратичного программирования. Это уже чисто вычислительная задача.

Этот вариант алгоритма называют алгоритмом с мягким зазором (soft-margin SVM), тогда как в линейно разделимом случае говорят о жестком зазоре (hard-margin SVM).

Преимущества метода:

- один из наиболее качественных методов;
- возможность работы с небольшим набором данных для обучения;
- сводимость к задаче выпуклой оптимизации, имеющей единственное решение.

Недостатки метода: сложная интерпретируемость параметров алгоритма и неустойчивость по отношению к выбросам в исходных данных.

Логистическая регрессия (logit model, logistic regression) является линейным методом классификации. Этот метод используется для предсказания вероятности возникновения некоторого события по значениям множества признаков. Для этого вводятся так называемая зависимая переменная u , которая может принимать лишь одно из двух значений – как правило, это числа 0 (событие не произошло) и 1 (событие произошло), и множество независимых переменных (также называемых признаками, предикторами или регрессорами) – вещественных x_1, \dots, x_n , на основе значений которых требуется вычислить вероятность принятия того или иного значения зависимой переменной. В случае классификации документов роль зависимой переменной выполняет категория c_i , а роль независимых переменных – набор документов d_1, \dots, d_n .

Для улучшения обобщающей способности алгоритма, то есть для уменьшения эффекта переобучения, на практике часто рассматривается логистическая регрессия с регуляризацией. Регуляризация заключается в том, что вектор параметров θ рассматривается как случайный вектор с некоторой заданной априорной плотностью распределения $p(\theta)$. Для обучения модели вместо метода наибольшего правдоподобия при этом используется метод максимизации апостериорной оценки, то есть должны быть найдены параметры θ , максимизирующие величину:
$$\prod_{i=1}^n P\{c_i | d_i, \theta\} \cdot p(\theta). \quad (13)$$

Мультиномиальная логистическая регрессия – это общий случай модели логистической регрессии, в которой зависимая переменная имеет более двух категорий. В модели мультиномиальной логистической регрессии для каждой категории зависимой переменной строится уравнение бинарной логистической регрессии. При этом одна из категорий зависимой переменной становится опорной, а все другие категории сравниваются с ней. Уравнение мультиномиальной логистической регрессии прогнозирует вероятность принадлежности к каждой категории зависимой переменной по значениям независимых переменных.

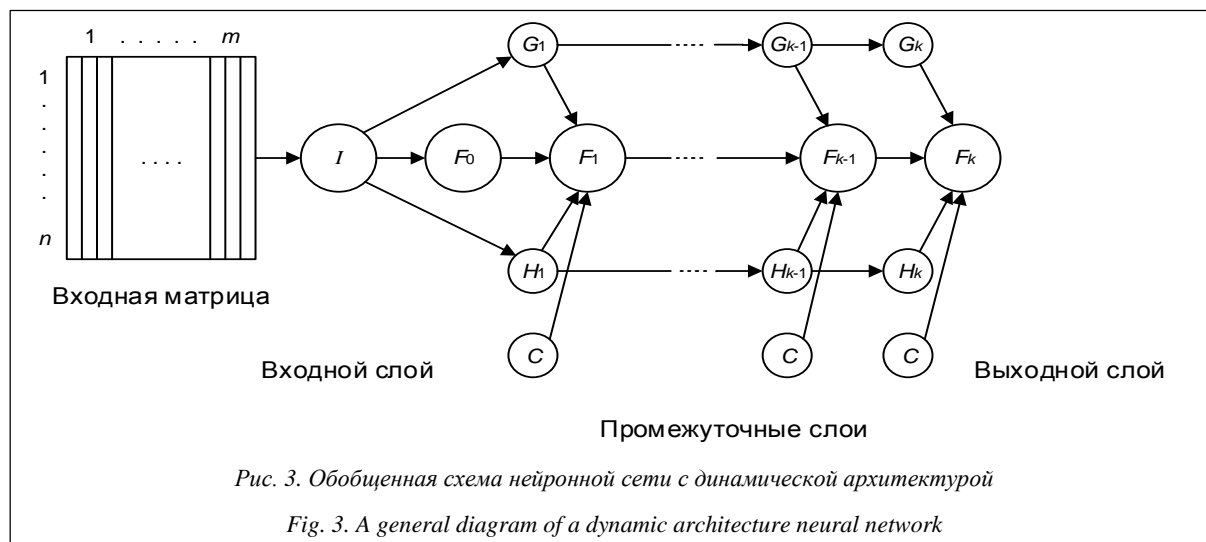
Вообще говоря, логистическую регрессию можно представить в виде однослойной нейронной сети с сигмоидальной функцией активации, веса которой – коэффициенты логистической регрессии, а вес поляризации – константа регрессионного уравнения:

$$P\{y = 1 | x\} = f(z). \quad (14)$$

Преимущества метода:

- является одним из наиболее качественных;
- поддерживает инкрементное обучение;
- имеет относительно простую программную реализацию алгоритма.

Недостатки метода: сложная интерпретируемость параметров алгоритма и неустойчивость по отношению к выбросам в исходных данных.



Методы на основе искусственных нейронных сетей. Существует большое количество разновидностей нейронных сетей, основные из них – сети прямого распространения, рекуррентные сети, радиально-базисные функции и самоорганизующиеся карты. Настройка весов может быть фиксированной или динамической.

В классических нейронных сетях прямого распространения (Feed Forward Back Propagation, FFBP) присутствуют входной слой, выходной слой и промежуточные слои: сигнал идет последовательно от входного слоя нейронов по промежуточным слоям к выходному. Примером такой структуры является многослойный перцептрон.

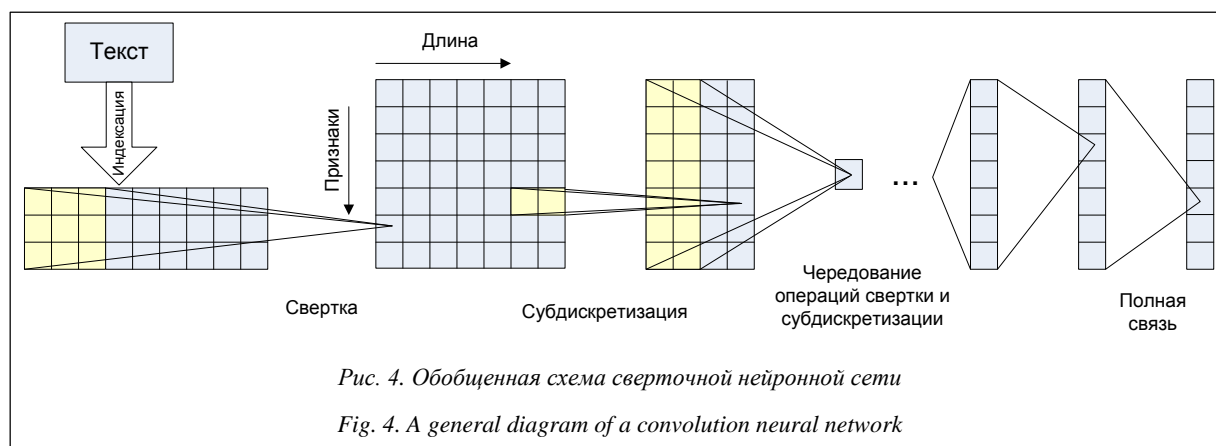
Для классификации документа d_i при помощи нейронной сети прямого распространения веса признаков документа подаются на соответствующие входы сети. Активация распространяется по сети; значения, получившиеся на выходах, и есть результат классификации. Стандартный метод обучения такой сети – метод обратного распространения ошибки. Суть его в следующем: если на одном из выходов для одного из обучающих документов получен неправильный ответ, то ошибка распространяется обратно по сети и веса ребер меняются так, чтобы уменьшить ошибку.

Количество промежуточных слоев нейронной сети может быть задано заранее, такую архитектуру называют динамической. В этом случае слои последовательно динамически генерируются до тех пор, пока не будет достигнут нужный уровень точности.

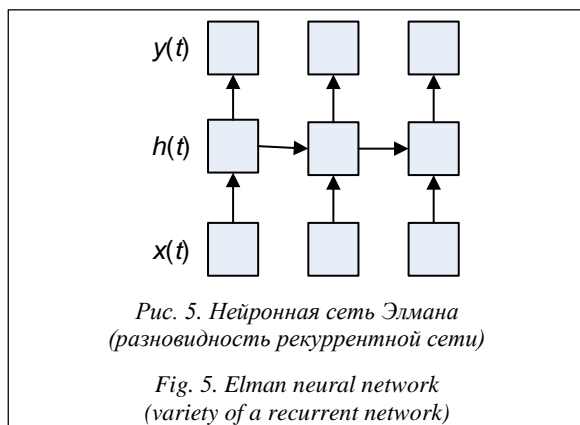
Обобщенная схема DAN2 приведена на рисунке 3, взятом из статьи [9]. Каждый элемент F_k представляет собой функцию, которая содержит текущий элемент накопленных знаний (Current Accumulated Knowledge Element), полученный на предыдущем шаге обучения сети. C обозначают константы. Вершины G_k и H_k представляют собой текущие остаточные нелинейные компоненты процесса по передаточной функции взвешенной и нормализованной суммы входных переменных (Current Residual Nonlinear Element).

Сверточная нейронная сеть – однонаправленная многослойная сеть с применением операции свертки, при которой каждый фрагмент входных данных умножается на матрицу (ядро) свертки поэлементно, а результат суммируется и записывается в аналогичную позицию выходных данных.

Обобщенная схема CNN приведена на рисунке 4, взятом из статьи [2].



Рекуррентная нейронная сеть получается из многослойного перцептрона введением обратных связей. Одна из широко распространенных разновидностей рекуррентных нейронных сетей – сеть Элмана – изображена на рисунке 5 [8]. В ней обратные связи идут не от выхода сети, а от выходов внутренних нейронов. Это позволяет учесть предысторию наблюдаемых процессов и накопить информацию для выработки правильной стратегии обучения. Главной особенностью рекуррентных нейронных сетей является запоминание последовательностей.



Скрытый слой $h(t)$ в период времени t вычисляется путем преобразования текущего входного слоя $x(t)$ и предыдущего скрытого слоя $h(t-1)$. Далее из скрытого слоя $h(t)$ результат поступает на выходной слой $y(t)$.

Преимущества метода:

- имеет очень высокое качество алгоритма при удачном подборе параметров;
- является универсальным аппроксиматором непрерывных функций;
- поддерживает инкрементное обучение.

Недостатки метода:

- вероятность возможной расходимости или медленной сходимости, поскольку для настройки сети используются градиентные методы;
- необходимость очень большого объема данных для обучения, чтобы достичь высокой точности;
- низкая скорость обучения;
- сложная интерпретируемость параметров алгоритма.

Оценка качества классификации

Для обучения и оценки качества классификации, как уже отмечалось ранее, требуются обучающая и тестовая выборки: $\Omega = L \cup T$. Прежде всего нужно выбрать обучающую и тестовую выборки, далее по обучающей выборке найти оптимальные признаки, а потом проверять качество на тестовой. Если сначала найти оптимальные признаки по всей выборке, а потом оценивать качество алгоритма, то

отобранные признаки уже оптимизируют качество, поэтому оценка будет слишком оптимистичной. Чтобы оценка качества классификатора была объективной, необходимо правильно выбрать соотношение объемов этих выборок. Если взять очень маленькую обучающую выборку, оценка качества будет слишком пессимистичной. Если тестовая выборка будет маленькая, оценка окажется неточной. Как правило, обучающую и тестовую выборки берут исходя из соотношения 70/30.

Однако есть более объективный способ оценки качества классификатора – кросс-валидация. Суть ее состоит в следующем: все множество Ω разбивается на k частей, каждая из них по очереди выступает как тестовая. Здесь важно сделать оптимальный выбор k . Обычно предпочитают брать $k = 5$ или $k = 10$. Главный недостаток такого способа оценки – большие трудозатраты.

Основным критерием при оценке качества классификации является комбинация точности и полноты.

Точность (precision) классификации в пределах класса – это доля найденных классификатором документов, действительно принадлежащих данному классу, относительно всех документов, которые система отнесла к этому классу.

Полнота (recall) классификации – это доля найденных классификатором документов, действительно принадлежащих классу, относительно всех документов этого класса в тестовой выборке.

Оценка качества работы классификатора производится на тестовой выборке. Вместе с тем работу системы оценивает эксперт (см. табл. 1).

Таблица 1

Оценка качества работы классификатора

Table 1

Classification quality assessment

Класс c_i		Экспертная оценка	
		Положительная	Отрицательная
Оценка системы	Положительная	TP	FP
	Отрицательная	FN	TN

В таблице приняты следующие условные обозначения: TP – истинно положительное решение; TN – истинно отрицательное решение; FP – ложно положительное решение; FN – ложно отрицательное решение.

Согласно определению, точность вычисляется следующим образом:

$$p = TP / (TP + FP). \quad (15)$$

Полнота вычисляется по формуле

$$r = TP / (TP + FN). \quad (16)$$

F-мера – характеристика качества работы алгоритма, которая объединяет в себе информацию о точности и полноте:

$$F_\beta = \frac{(\beta^2 + 1)pr}{\beta^2 \cdot p + r}, \quad (17)$$

где $0 \leq \beta < \infty$.

При $0 \leq \beta < 1$ большее значение имеет точность.

При $\beta = 1$ точность и полнота равноправны, тогда $F_\beta = 2pr / (p + r)$.

При $1 < \beta < \infty$ большее значение имеет полнота.

Часто можно встретить другую формулу для вычисления точности (*accuracy*). Эту величину иногда называют правильностью или аккуратностью метода:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} . \quad (18)$$

В некоторых случаях удобнее от долей перейти к процентам, умножив полученную величину на 100.

Иногда для сравнения алгоритмов классификации используют специфические характеристики, такие как точка безубыточности, или сбалансированная точность.

Точка безубыточности (*break even point*, BEP) – величина, заимствованная из экономики, отражающая объем производства и реализации продукции, при котором расходы будут компенсированы доходами, а при производстве и реализации каждой последующей единицы продукции предприятие начинает получать прибыль. В контексте рассматриваемой задачи точка безубыточности используется как мера качества классификации. Точка безубыточности наравне с F-мерой является сбалансированной характеристикой точности и полноты. Более подробное пояснение можно найти в [11].

Под быстродействием классификатора понимается время, затрачиваемое на отнесение документа к одному из классов. Применительно к задачам классификации текстов быстродействие измеряется как процессорное время (в секундах) или как количество вычислительных операций, необходимое для классификации. Измерение производят на обучающей выборке для оценки скорости процесса обучения и отдельно на тестовой выборке. Следует отметить, что высокие затраты при обучении в дальнейшем оправдываются за счет многократного использования настроенного классификатора.

Ясно, что увеличение точности классификации обычно приводит к снижению быстродействия из-за усложнения решающего правила, используемого в алгоритме классификации, а увеличение быстродействия сопровождается понижением точности из-за упрощения работы классификатора.

Эксперименты по сравнению методов

В работе [9] предложен алгоритм классификации на основе нейронных сетей с динамической архитектурой DAN2. Этот вид сетей был выбран на основании экспериментального сравнения DAN2 с обычными нейронными сетями прямого распространения (FFBP) и рекуррентными нейронными сетями (RNN). Для сравнения качества классификации в [9] были рассмотрены DAN2, KNN и SVM.

Классификация проводилась на широко известной коллекции данных Reuters-21578 (<http://www.daviddlewis.com/resources/testcollections/reuters21578/>), собранной и размеченной в 2004 году Д. Льюисом. Коллекция содержит 21 578 документов из ленты новостей Reuters. Обучающая выборка состоит из 9 603 документов. Тестовая выборка включает в себя 3 299 документов. В эксперименте разбиение производилось на десять наиболее часто встречающихся категорий, связанных с экономикой (нефть, пшеница, кукуруза, торговля, деньги и пр.). Авторы пришли к выводу, что оптимальное количество признаков для каждого из методов – около 2 200. Сравнение качества алгоритмов осуществлялось при помощи точности, полноты, точки безубыточности и F-меры. Установлено, что в зависимости от категории BEP для DAN2 – это 83,17–99,23 %, для KNN – 74,00–97,30 %, для SVM – 75,00–98,50 %. Точность DAN2 для разных классов варьируется от 97,56 до 100 %, полнота – 68,52–99,66 %, F-мера – 80,63–99,23 %.

Время, потраченное на обучение классификатора с использованием DAN2, варьируется в зависимости от категории – 4,078–410,2969 с, время работы уже обученного классификатора на тестовой выборке для выбранных десяти категорий составляет 0,0081–9,5859 с. Для экспериментов использовался многоядерный сервер со следующей конфигурацией: 2 Intel Quad Core Xeon @ 3,2 GHz, 16 GB of RAM, Adaptec Raid Controller with 4 SAS hard drives in RAID 1/0 configurations. Операционная система SuSE Linux Enterprise Server (SLES, 11) 64-bit. Еще пробовали VMWare Server, OpenMPI.

На основе полученных экспериментальных данных можно прийти к выводу, что DAN2 опережает KNN для всех десяти категорий и опережает SVM для девяти из десяти категорий. Вместе с тем следует отметить, что применение нейронных сетей сильно замедляет работу классификатора на этапе обучения.

В работе [4] утверждается, что формально SVM и нейронная сеть прямого распространения (FFNN) имеют похожую структуру, так как выходная функция может быть представлена в виде линейной комбинации простых функций, то есть

$$f(x) = b + \sum_{k=1}^M \lambda_k h(w_k, x) . \quad (19)$$

В таком случае количество скрытых нейронов является долей числа опорных векторов (табл. 2). Более подробное описание используемой в таблице общепринятой терминологии можно найти, например, в [12].

Кроме того, SVM используется в задаче выпуклой оптимизации, которая всегда позволяет найти глобальный минимум и единственное решение, в то время как FFNN тренируется при помощи метода градиентного спуска, который не всегда сходится к оптимальному (глобальному) решению. В статье [4] предложены техники для минимизации

случая локальной сходимости, а также показано, что масштабированный метод сопряженных градиентов не сходится реже, чем традиционный метод сопряженных градиентов или метод обратного распространения с использованием градиентного спуска.

Таблица 2

Соотношение элементов SVM и FFNN
Table 2
Correlation of SVM and FFNN

Элемент метода	SVM	FFNN
M	Количество опорных векторов	Количество вершин в скрытом слое
h	Функция-ядро	Функция активации
$\{w_k\}_{k=1}^M$	Опорные векторы	Веса скрытого слоя
$\{\lambda_k\}_{k=1}^M$	Коэффициенты задачи выпуклой оптимизации	Веса выходного слоя

В подтверждение своих наблюдений авторы приводят описание эксперимента по разделению отзывов о фильмах, книгах, GPS и фотоаппаратах на положительные и отрицательные с использованием методов SVM, NB и ANN. В методе SVM в качестве обычного нелинейного ядра была взята радиальная базисная функция. В методе с искусственными нейронными сетями было отдано предпочтение прямооточной нейронной сети (однонаправленной сети с одним скрытым слоем). Для обучения нейронной сети использовался алгоритм обратного распространения ошибки (Backpropagation). Чтобы ускорить процесс обучения и сократить риск переобучения, применялась технология «ранней остановки».

Результаты классификации сравнивались на сбалансированных и несбалансированных данных. Данные для категории «фильмы» были взяты из популярной, часто цитируемой базы Movie Review Data (<http://www.cs.cornell.edu/people/pabo/movie-review-data/>), для остальных категорий авторы собирали коллекции самостоятельно с сайта Amazon (<http://www.amazon.com/>) по 2 000 отзывов для каждого класса.

Характеристиками для сравнения являлись точность (precision), полнота (recall), аккуратность (accuracy) и время в секундах. Сравнение осуществлялось с помощью 10-проходной кросс-валидации. Количество признаков в экспериментах варьировалось от 50 до 5 000. В среднем наилучшие результаты были получены при 500–1 000. Было замечено, что на 5 000 терминов для ANN время на обучение значительно увеличивается (с 3,70 с до 69,40 с), а время работы не меняется; для SVM, наоборот, время на обучение не меняется, но слишком большое количество признаков сильно сказывается на длительности работы.

На сбалансированных данных было проведено 28 тестов для четырех категорий. ANN показал

лучший результат, чем SVM, в 13 тестах (t-тест с $p < 0,05$); SVM превзошел ANN только в 2, хотя в целом разница в результатах не превысила 3 %.

Худшие результаты получены для класса «книги». Точность 0,88 % при 3 000 терминов для SVM; 0,86 % при 3 000 и 4 000 терминов для SVM. Полнота 0,8 при 1 000 терминов для ANN; 0,88 при 3 000 для SVM. Лучшая аккуратность (accuracy) 81,8 % при 1 000 терминов достигается для ANN.

Для остальных трех категорий результаты лучше. Для класса GPS лучшая точность 0,96 и лучшая полнота 0,99 достигаются методом NB, лучшая аккуратность (accuracy) 87,3 % – методом ANN. Для класса «фильмы» лучшая точность 0,95 достигается NB, на втором месте 0,87 – ANN, лучшая полнота 0,98 – NB, на втором месте 0,87 – ANN, лучшая аккуратность (accuracy) 86,5 % получена методом ANN. Для класса «фотоаппараты» лучшая точность 0,94 получена методом NB, лучшая полнота 0,96 – NB, лучшая аккуратность (accuracy) 90,3 % – ANN.

Время работы зависит от количества векторов для SVM и количества слоев для ANN. Наравне с ними рассматривался Байес. Он, бесспорно, быстрее всех (0,01–0,02 с) при любом количестве терминов для любого класса и в некоторых случаях, как ни странно, показывал лучшую точность. Время на обучение для класса «книги»: SVM – 0,22–1,5 с, ANN – 3,7–69,4 с в зависимости от количества признаков (50–5 000). Время на обучение для класса GPS: SVM – 0,2–1,3 с, ANN – 3,1–75,4 с. Время на обучение для класса «фильмы»: SVM – 0,27–5,6 с, ANN – 2,3–65,5 с. Время на обучение для класса «фотоаппараты»: SVM – 0,2–1,1 с, ANN – 4,5–77,2 с. К сожалению, в статье отсутствуют данные об аппаратном обеспечении, на котором проводилось исследование.

В работе [10] рассматриваются методы обнаружения ложных высказываний в текстах, когда люди намеренно говорят неправду, пытаясь обмануть. Авторы исследовали высказывания людей, вовлеченных в преступления на военных базах. Подозреваемые и свидетели описывали события своими словами. Сотрудники правоохранительных органов находили в архивных данных либо подтверждения, либо опровержения этим высказываниям. Таким образом оценивали истинность высказываний либо их ложность. Проанализировано 371 сообщение из специальных архивов о различных видах преступлений: дорожные нарушения, магазинные кражи, нападения и поджоги. Большой проблемой было собрать такую коллекцию данных, для которой можно установить истинность/ложность высказываний.

Для классификации первоначально экспертами был составлен перечень из 31 признака. Далее при помощи критерия хи-квадрат было выбрано 13 наиболее подходящих признаков: количество глаголов движения, личных местоимений, количество

слов с оттенком намерения, причины, с указанием времени, лексическое разнообразие и пр. Некоторые из отобранных признаков весьма специфичны и требуют составления семантических словарей.

В эксперименте проводилось сравнение многослойного перцептрона (MLP, разновидность FFNN), модификации деревьев решений (CART), логистической регрессии и ансамбля классификаторов. Для построения ансамбля классификаторов, как правило, используются два основных метода: бустинг (boosting) и бэггинг (bagging). При бустинге происходит последовательное обучение классификаторов. Например, первый классификатор обучается на всем наборе данных, второй – на выборке примеров, а третий – на наборе тех данных, в которых результаты первых двух классификаторов разошлись. Бэггинг использует параллельное обучение базовых классификаторов, то есть бэггинг является улучшающим объединением, а бустинг – улучшающим пересечением.

Для проверки использовалась 10-проходная кросс-валидация. Для метода MLP достигнута точность 73,46 %, для CART – 71,60 %, для логистической регрессии – 67,28 %. В результате был сделан вывод, что наиболее высокая точность, 74,07 %, достигается на ансамбле классификаторов. Преимуществом ансамблевых классификаторов является качество их работы на неравномерно распределенных данных. Эта особенность важна при потоковой обработке данных, когда некоторые редкие классы документов, появляющиеся и исчезающие в потоке, порой непросто обнаружить.

В работе [8] рассмотрено решение сразу двух задач: извлечение терминов и классификация для англоязычного и русскоязычного корпусов. Необходимо было определить эмоциональную окраску отзывов о ресторанах и автомобилях, то есть разделить отзывы на положительные и отрицательные, построив бинарный классификатор для каждой из категорий «рестораны» и «автомобили».

Для классификации использовались рекуррентные нейронные сети, в частности, нейронные сети Элмана (простые и двунаправленные BRNN) и LSTM. Для извлечения аспектных терминов в сравнении участвовали несколько методов: два вида многослойного перцептрона (MLP), логистическая регрессия и условные случайные поля (CRF). Для условных случайных полей в качестве признаков использовались основы слов и принадлежность частям речи (проводилась процедура POS-tagging). Использовались две метрики для извлечения аспектных терминов: на основе точного количества и на основе пропорционального перекрытия.

Сравнение результатов классификации для английских текстов проводилось на недавно собранной коллекции SemEval-2014 ABSA Restaurants (<http://metashare.islp.gr:8080/repository/search/?q=SemEval-2014+ABSA+Restaurants>). Обучающая выборка состояла из 3 041 сообщения, тестовая –

из 800. Для проверки качества методов на русских текстах была использована коллекция отзывов о ресторанах и автомобилях, собранная для проведения соревнований SentiRuEval-2015 в рамках конференции «Диалог». Для проверки полученных результатов применялась 7-проходная кросс-валидация.

Лучшие результаты как при извлечении терминов, так и при классификации отзывов показал метод LSTM. При извлечении терминов F-мера для LSTM составила 79,80 %. При классификации точность – 69,70 %. Метод LSTM для русскоязычных данных для класса «рестораны» показал точность 61,1 %, F-меру – 70,2 %. Для класса «автомобили» результаты хуже: точность – 58,0 %, F-мера – 62,4 %.

Проблема тематической классификации коротких текстовых сообщений (от нескольких слов до 2 предложений), например, смс-сообщений, комментариев к новостям, на форумах, в социальных сетях, рассматривается в [7]. Основная сложность, возникающая при решении этой проблемы, – определение набора признаков, по которым предстоит классифицировать. В качестве признаков классификации принято рассматривать слова и словосочетания, буквы и буквосочетания. Один из недостатков распространенных на сегодняшний день методов – привязка к конкретному естественному языку и опора на словари. В данной публикации уделяется внимание определению универсальных методов и дифференцирующих признаков для текстов на различных естественных языках.

В данной работе также описан алгоритм построения и обучения классификатора на основе метода взаимной информации (PMI). Применялась процедура POS-tagging, и для классификации были выбраны следующие признаки: *N* – существительные, *NA* – существительные и прилагательные, *NAV* – существительные, прилагательные, глаголы, *NNP* – существительные и именные группы, *VVP* – глаголы и глагольные группы, *Stem* – псевдоосновы словоупотреблений текста, полученные алгоритмами аналитического морфологического анализа (имеются в виду широко известные алгоритмы Портера, Ловинса, Пейса–Хаска и пр.).

Принадлежность текста к категории определяется наличием в нем признаков, релевантных данной категории и коррелирующих с признаками рассматриваемой категории, а также отсутствием нерелевантных признаков и признаков, не коррелирующих с признаками данной категории [7]. При таком подходе тексту можно сопоставить информационную матрицу I , элементы которой определяются как пара: $I_{ij} = \{\rho_{ij}, \varepsilon_{ij}\}$, где ρ_{ij} – коэффициент релевантности; ε_{ij} – коэффициент корреляции.

Для коэффициентов релевантности и корреляции справедливо следующее утверждение: большие значения этих коэффициентов соответствуют признакам, наиболее точно характеризующим

выбранный класс. Пороговые значения релевантности и корреляции служат параметрами, определяющими точность. В процессе классификации вычисляются коэффициенты релевантности и корреляции для каждого текста как суммы соответствующих коэффициентов для данного класса по всем вхождениям признаков. Документ считается отнесенным к тем категориям, для которых произошло превышение пороговых значений по обеим характеристикам: как по коэффициенту корреляции, так и по коэффициенту релевантности. Пороговые значения для каждой категории могут быть заданы пользователем или же рассчитаны автоматически по обучающей выборке.

В работе [7] представлены результаты экспериментов для метода NB и метода на основе PMI. Обучение классификаторов проводилось на созданных экспертами выборках текстов с сайтов из Интернета: для русского языка объемом 57,3 Мб, башкирского – 1,87 Мб, татарского – 2,68 Мб. В качестве классов условно были выбраны «наркотики», «насилие», «национализм», «отрицание традиционных ценностей», «порнография», «терроризм», «фашизм», «экстремизм».

Учет выбранных морфологических признаков оказывает различное влияние на качество классификации в зависимости от класса. Для некоторых классов (например «фашизм») могут оказывать положительное влияние существительные и именные группы, а на определение некоторых тематик (например «наркотики», «фашизм») отрицательное влияние оказывает учет глагольных групп.

Лучшие значения F-меры достигаются предложенным методом (на основе PMI, в качестве признаков – псевдоосновы) для классов: «жестокость» – 0,913, «отрицание традиционных ценностей» – 0,862, «наркотики» – 0,765. В итоге можно сделать вывод, что псевдоосновы, выделенные аналитическим алгоритмом морфологического анализа, могут считаться универсальными дифференцирующими признаками при классификации коротких текстовых сообщений.

В работе [2] предложен алгоритм классификации с применением сверточных нейронных сетей (CNN), проведено сравнение этого метода с методом на основе рекуррентных нейронных сетей (LSTM) и мультиномиальной логистической регрессией (logit model) в разных вариациях («мешок слов», TF-IDF, n -граммы, Word2vec). Тестирование сверточных нейронных сетей проводилось на данных, основанных на словах (word-based) и на символах (character-based).

Особенность метода CNN заключается в необходимости использования очень больших коллекций для обучения. Большинство открытых коллекций для классификации текстов (даже на английском языке) слишком малы, поэтому для эксперимента авторы самостоятельно собрали тексты с новостных сайтов, обзоры и отзывы пользо-

вателей, данные из DBPedia, которая содержит структурированные данные из википедии. В результате были получены коллекции данных на английском и китайском языках. Объемы собранных коллекций приведены в таблице 3: обучающие выборки содержали от 120 000 до 3 600 000 текстов, тестовые – от 7 600 до 650 000 текстов.

Таблица 3

Количество классов и объем обучающей и тестовой выборки для реализации метода CNN

Table 3

A number of classes, training and test set volume for CNN method implementation

Название коллекции	Количество классов	Обучающая выборка	Тестовая выборка
AG's News	4	120 000	7 600
Sogou News	5	450 000	60 000
DBPedia	14	560 000	70 000
Yelp Review Polarity	2	560 000	38 000
Yelp Review Full	5	650 000	50 000
Yahoo! Answers	10	1 400 000	60 000
Amazon Review Full	5	3 000 000	650 000
Amazon Review Polarity	2	3 600 000	400 000

Наименьшие погрешности измерения 1,31–7,64 % были достигнуты для моделей, использующих в качестве признаков n -граммы, а также сверточные нейронные сети (погрешность измерения 4,93–40,43 %) в зависимости от коллекции.

Самые плохие результаты показала модель с применением метода Word2vec. Это означает, что получивший широкое распространение метод представления слов Word2vec в виде векторов не дает преимуществ в задаче классификации текстов. Хотя авторам статьи [2] еще предстоит детально интерпретировать полученные результаты, а также продолжать эксперименты на коллекциях текстов для других языков, сейчас можно сделать вывод, что для задачи классификации текстов лучшим оказался символьный уровень, когда рассматриваются буквосочетания (без привязки к конкретному языку).

Результаты исследования

Классификация текстов является одной из основных задач компьютерной лингвистики, поскольку к ней сводится ряд других задач: определение тематической принадлежности текстов, автора текста, эмоциональной окраски высказываний и др.

Формально задачу классификации текстов можно описать следующим образом. Имеется мно-

жество документов и множество возможных категорий (классов). Требуется построить классификатор, относящий выбранный документ к одной из нескольких заранее определенных категорий на основании содержания документа. Наиболее распространенный современный подход к классификации основывается на методах машинного обучения. Согласно этим методам, набор правил или критериев принятия решения текстового классификатора вычисляется автоматически на основе обучающих данных. Обучающими данными являются образцы документов из каждого класса.

Решение задачи классификации состоит из четырех последовательных этапов: предобработка и индексация документов, уменьшение размерности пространства признаков, построение и обучение классификатора с помощью методов машинного обучения, оценка качества классификации.

Для предварительной обработки и индексации документа (то есть при построении некоторой числовой модели текста) обычно применяется одна из трех моделей: модель «мешка слов», Word2vec и модель, основанная на учете n -грамм. Для реализации первой и второй моделей необходимы дополнительные знания о морфологической и синтаксической структуре языка. Применение символьных n -грамм позволяет не накладывать ограничения на использование конкретного языка, поэтому в ряде случаев является предпочтительным.

Вычислительная сложность различных методов классификации напрямую зависит от размерности пространства признаков. За счет уменьшения размерности пространства терминов можно снизить эффект переобучения – явление, при котором классификатор ориентируется на случайные или ошибочные характеристики обучающих данных, а не на важные и значимые. Переобученный классификатор хорошо работает на тех экземплярах, на которых он обучался, и значительно хуже на тестовых данных. Чтобы избежать переобучения, количество обучающих примеров должно быть соразмерно числу используемых терминов, поэтому для эффективной работы классификатора часто прибегают к сокращению числа используемых признаков (терминов). Для уменьшения размерности пространства терминов применяют такие методы, как LSA, TF-IDF, PMI, CRF, IG. Наибольшее распространение из них получил метод TF-IDF.

Выводы

В статье были рассмотрены следующие наиболее распространенные методы построения и обучения классификатора: NB, KNN, SVM, DT, логистическая регрессия и алгоритмы глубокого обучения, основанные на искусственных нейронных сетях (FFBP, RNN, DAN2, CNN).

Для обучения и оценки качества классификации необходимо подготовить обучающую и тестовую

выборки, далее по обучающей выборке найти оптимальные признаки, а затем проверять качество на тестовой выборке. Чтобы оценка качества классификатора была объективной, требуется правильно выбрать соотношение объемов этих выборок. Как правило, обучающую и тестовую выборки берут исходя из соотношения 70/30. Более объективным способом оценки качества классификатора является кросс-валидация.

Общепризнанными характеристиками качества работы классификатора являются точность, полнота и их комбинация (F-мера). На основе проведенного исследования можно сделать вывод, что наилучшее соотношение этих характеристик достигается при использовании методов SVM (точность 80–85 %, полнота 83–87 %) и CNN (точность 90–95 %, полнота 80–85 %). Помимо характеристик качества классификации, целесообразно учитывать также другие факторы: время работы алгоритма, возможность работы алгоритма в инкрементном режиме, количество предварительной информации, необходимой для классификации, независимость от языка. Скорость работы алгоритма NB одна из самых высоких, однако точность для различных экспериментов сильно варьируется (71–90 %). При потоковой обработке текстов классификация документов должна осуществляться одновременно с поступлением их из источника, поэтому предпочтение должно отдаваться инкрементным алгоритмам, таким как CNN или SVM.

В настоящее время по-прежнему остается нерешенным вопрос определения набора классифицирующих признаков, их количества и способов вычисления весов. При выборе определенного метода следует помнить, что при большом количестве признаков (около 5 000) время на обучение для нейронных сетей значительно увеличивается, а время работы не меняется; для SVM, наоборот, время на обучение не меняется, но слишком большое количество признаков сильно сказывается на длительности работы. Оптимальным является 500–1 000 признаков, в некоторых случаях – до 2 200 признаков. В качестве признаков удобно рассматривать частоты символьных n -грамм, чтобы не накладывать ограничения на использование конкретного языка.

В алгоритмах глубокого обучения точность классификации существенно зависит от наличия обучающей выборки подходящего размера. Подготовка такой выборки – очень трудоемкий процесс. Подбор параметров некоторых алгоритмов на этапе обучения до сих пор также остается открытой проблемой. Согласно результатам проведенного исследования, для обучения и тестирования классификатора с использованием метода SVM на русском языке нужна размеченная коллекция текстов объемом 1 000–2 000 текстов для достижения точности 80–85 %. Для обучения и тестирования классификатора с применением метода CNN необхо-

димо собрать и подготовить коллекцию текстов на русском языке объемом около 1 000 000 документов для достижения точности 90–95 %.

Следует отметить, что большинство упоминаемых в обзоре экспериментов проводилось на коллекциях англоязычных текстов. Довольно часто встречаются статьи с описанием исследований применительно к китайскому языку. Исследования по сравнению различных методов классификации для русскоязычных текстов проводятся в основном в контексте задачи сентимент-анализа, где рассматриваются два класса: положительный и отрицательный.

Создание общедоступной коллекции необходимого размера позволило бы российским исследователям активнее изучать проблемы автоматической обработки текстовой информации в целом и вместе с тем разрабатывать новые инструменты для решения прикладных задач в данной области.

Работа выполнена при финансовой поддержке Минобрнауки РФ (договор № 02.G25.31.0146) в рамках реализации Постановления Правительства РФ № 218.

Литература

1. Aggarwal C. Data classification: algorithms and applications. CRC Press, 2014, chap. 9, pp. 245–273.
2. Xiang Zhang, Junbo Zhao, Yann LeCun. Character-level convolutional networks for text classification. Proc. Neural Inform. Processing Systems Conf. (NIPS 2015). Montreal, Canada, 2015. URL: <https://arxiv.org/abs/1509.01626> (дата обращения: 18.07.2016).
3. Ju R. et al. An Efficient Method for Document Categorization Based on Word2vec and Latent Semantic Analysis. 2015 IEEE Intern. Conf. on Comp. and Inform. Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing. Liverpool, UK, 2015, pp. 2276–2283.
4. Moraes R., Valiati J.F., and Gavião Neto W.P. Document-level sentiment classification: An empirical comparison between SVM and ANN. Expert Systems with Applications, 2013, no. 40, pp. 621–633.
5. Pontiki M Galanis D., Pavlopoulos J., Papageorgiou H., Androutsopoulos I., Manandhar S. SemEval-2014 Task 4: Aspect based sentiment analysis. The 8th Intern. Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland. 2014, pp. 27–35.
6. Medhat W., Hassan A., Korashy H. Sentiment analysis algorithms and applications: a survey. Ain Shams Eng. Jour., 2014, no. 5, pp. 1093–1113.
7. Поляков И.В., Соколова Т.В., Чеповский А.А., Чеповский А.М. Проблема классификации текстов и дифференцирующие признаки // Вестн. НГУ. Сер.: Информационные технологии. 2015. Т. 13. Вып. 2. С. 55–63.
8. Tarasov D.S. Deep recurrent neural networks for multiple language aspect-based sentiment analysis. Computational Linguistics and Intellectual Technologies. In Proc. Annual Intern. Conf. «Dialogue-2015». Moscow, 2015, vol. 2, iss. 14 (21), pp. 65–74.
9. Ghiassi M., Olschmke M., Moon B., Arnaudo P. Automated text classification using a dynamic artificial neural network model. Expert Syst. with Applications, 2012, no. 39, pp. 10967–10976.
10. Fuller C.M., Biros D.P. and Delen D. An investigation of data and text mining methods for real world deception detection. Expert Syst. with Applications, 2011, no. 38, pp. 8392–8398.
11. Yang Y. An evaluation of statistical approaches to text categorization. Information Retrieval Jour., 1999, vol. 1, iss. 1, pp. 69–90.
12. Haykin S. Neural networks: A comprehensive foundation (2nd ed.). Pearson Education, Singapore, 2001, 824 p.

Software & Systems

DOI: 10.15827/0236-235X.117.085-099

Received 19.07.16

2017, vol. 30, no. 1, pp. 85–99

AUTOMATIC TEXT CLASSIFICATION METHODS

T.V. Batura^{1,2}, Ph.D. (Physics and Mathematics), Leading Researcher, Senior Researcher, tatiana.v.batura@gmail.com

¹ Novosibirsk State University, Pirogov St. 2, Novosibirsk, 630090, Russian Federation

² A.P. Ershov Institute of Informatics Systems (IIS), Siberian Branch of the Russian Federation Academy of Sciences, Lavrentev Av. 6, Novosibirsk, 630090, Russian Federation

Abstract. Text classification is one of the main tasks of computer linguistics because it unites a number of other problems: theme identification, authorship identification, sentiment analysis, etc. Content analysis in telecommunication networks is of great importance to ensure information security and public safety. Texts may contain illegal information (including data related to terrorism, drug trafficking, organization of protest movements and mass riots). This article provides a survey of text classification methods. The purpose of this survey is to compare modern methods for solving the text classification problem, detect a trend direction, and select the best algorithm for using in research and commercial problems.

A well-known modern approach to text classification is based on machine learning methods. It should take into account the characteristics of each algorithm for selecting a particular classification method. This article describes the most popular algorithms, experiments carried out with them, and the results of these experiments. The survey was prepared on the basis of scientific publications which are publicly available on the Internet, made in the period of 2011–2016, and highly regarded by the scientific community.

The article contains an analysis and a comparison of different classification methods with the following characteristics: precision, recall, running time, the possibility of the algorithm in incremental mode, amount of preliminary information necessary for classification, language independence.

Keywords: text classification, analysis of text information, data mining, text mining, natural language processing, classification quality, machine learning, deep learning, neural networks.

Acknowledgements. The work has been financially supported by Ministry of Education and Science of the Russian Federation (contract no. 02.G25.31.0146) as a part of RF Government Regulation no. 218 execution.

References

1. Aggarwal C. *Data Classification: Algorithms and Applications*. CRC Press, 2014, pp. 245–273.
2. Zhang X., Zhao J., LeCun Y. Character-level Convolutional Networks for Text Classification. *Proc. of the Neural Information Processing Systems Conf. (NIPS 2015)*. Montreal, Canada, 2015. Available at: <https://arxiv.org/abs/1509.01626> (accessed July 18, 2016).
3. Ju R. An Efficient Method for Document Categorization Based on Word2vec and Latent Semantic Analysis. *2015 IEEE Int. Conf. on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Automatic and Secure Computing; Pervasive Intelligence and Computing*. Liverpool, UK, 2015, pp. 2276–2283.
4. Moraes R., Valiati J.F., Gavião Neto W.P. Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*. 2013, no. 40, pp. 621–633.
5. Pontiki M., Galanis D., Pavlopoulos J., Papageorgiou H., Androutsopoulos I., Manandhar S. SemEval-2014 Task 4: Aspect based sentiment analysis. *Proc. 8th Int. Workshop on Semantic Evaluation (SemEval 2014)*. Dublin, Ireland, 2014, pp. 27–35.
6. Medhat W., Hassan A., Korashy H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journ.* 2014, no. 5, pp. 1093–1113.
7. Polyakov I.V., Sokolova T.V., Chepovsky A.A., Chepovsky A.M. Text classification problem and features set. *Vestn. NGU. Ser.: Informatsionnye tekhnologii* [Novosibirsk State Univ. Journ. of Information Technologies]. 2015, vol. 13, iss. 2, pp. 55–63 (in Russ.).
8. Tarasov D.S. Deep Recurrent Neural Networks for Multiple Language Aspect-Based Sentiment Analysis. *Computational Linguistics and Intellectual Technologies: Proc. of Annual Int. Conf. "Dialogue-2015"*. Moscow, Russia, 2015, vol. 2, iss. 14 (21), pp. 65–74.
9. Ghiassi M., Olschmke M., Moon B., Arnaudo P. Automated text classification using a dynamic artificial neural network model. *Expert Systems with Applications*. 2012, no. 39, pp. 10967–10976.
10. Fuller C.M., Biros D.P. and Delen D. An investigation of data and text mining methods for real world deception detection. *Expert Systems with Applications*. 2011, no. 38, pp. 8392–8398.
11. Yang Y. An evaluation of statistical approaches to text categorization. *Information Retrieval Jour.* 1999, vol. 1, iss. 1, pp. 69–90.
12. Haykin S. *Neural networks: A comprehensive foundation*. 2nd ed., Pearson Education Publ., Singapore, 2001, 824 p.

Примеры библиографического описания статьи

1. Батура Т.В. Методы автоматической классификации текстов // Программные продукты и системы. 2017. Т. 30. № 1. С. 85–99; DOI: 10.15827/0236-235X.117.085-099.
2. Batura T.V. Automatic text classification methods. *Programmnye produkty i sistemy* [Software & Systems]. 2017, vol. 30, no. 1, pp. 85–99 (in Russ.); DOI: 10.15827/0236-235X.117.085-099.

УДК 004.021

DOI: 10.15827/0236-235X.117.100-105

Дата подачи статьи: 30.01.17

2017. Т. 30. № 1. С. 100–105

МЕТОДЫ АВТОМАТИЗИРОВАННОГО АНАЛИЗА КОРОТКИХ НЕСТРУКТУРИРОВАННЫХ ТЕКСТОВЫХ ДОКУМЕНТОВ

П.Ю. Козлов, аспирант, originaldod@gmail.com

*(Смоленский филиал Национального исследовательского университета МЭИ,
Энергетический проезд, 1, г. Смоленск, 214013, Россия)*

В работе рассматриваются задачи автоматизированного анализа текстовых документов в органах исполнительной и законодательной власти. Выделяется группа признаков для классификации текстовых документов, приводятся их типы, методы анализа и рубрицирования. Определяется перечень типов документов, которые необходимо классифицировать.

Для анализа коротких неструктурированных текстовых документов предлагается использовать метод классификации на основе весовых коэффициентов, экспертной информации, нечеткого логического вывода, для которого усовершенствована вероятностная математическая модель, разработан способ обучения и экспериментально подобрано соотношение весовых коэффициентов. Предварительно разработанный метод необходимо обучить.

На этапе обучения слова тезауруса для каждой предметной области разбиваются на три типа: уникальные, редкие и общие, и в зависимости от типа словам присваиваются весовые коэффициенты. Для поддержания актуальности весовых и частотных коэффициентов предлагается использовать динамическую кластеризацию. Разработанный метод позволяет анализировать описанные документы, а также учесть динамичность тезауруса рубрик.

Представлена схема работы системы автоматизированного анализа неструктурированных текстовых документов, написанных на естественном языке, различных типов: длинные, короткие, очень короткие. В зависимости от типа документа используется соответствующий метод анализа, который имеет наилучшие показатели точности и полноты при анализе текстовых документов данного типа. В качестве синтаксического анализатора используется парсер Malt-Parser, обученный на национальном наборе русского языка.

Результатом работы всей системы можно считать базу знаний, в которую попадают все извлеченные знания и их отношения. База знаний постоянно пополняется и используется работниками исполнительной и законодательной власти для обработки поступающих запросов.

Ключевые слова: автоматизированный анализ текстов, короткие неструктурированные тексты, динамичный тезаурус.

Органы исполнительной и законодательной власти становятся все более открытыми для населения, и это приводит к росту количества заявлений, жалоб и предложений. Значительная часть этих обращений поступают в электронном виде, что обуславливает необходимость их автоматизированной обработки. По каждому обращению необходимо выполнить следующие действия:

- принять обращение, определить суть проблемы и специалиста для решения данной проблемы;
- отправить обращение конкретному специалисту;
- принять некоторые меры по решению указанной проблемы;
- написать ответ с отчетом о проделанной работе.

Специфика автоматизированных систем обработки текстовых документов такого рода в нестационарности тезауруса ключевых слов, которые с выходом новых правовых документов кардинально изменяются, к тому же поступающие от населения обращения являются неструктурированными и короткими, что затрудняет статистический анализ.

Анализ литературных источников [1–5] показал, что текстовые документы можно классифицировать по четырем признакам: структурированность, объем, значимость ключевых слов и синтаксическая связность (рис. 1).

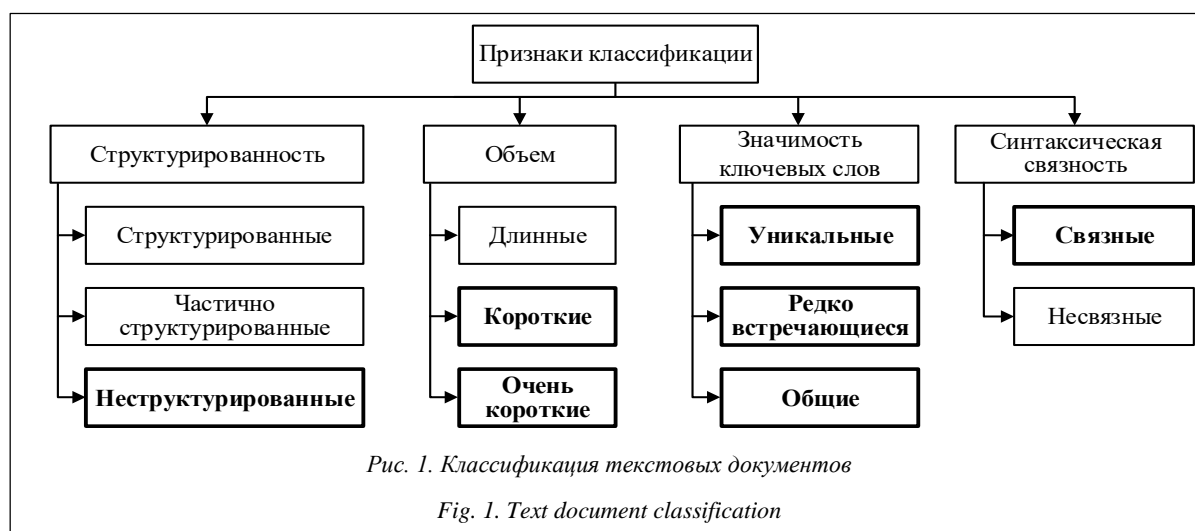
Поступающие в региональную систему обращения граждан обычно относятся к типу неструктурированных, коротких или очень коротких связных текстовых документов. Эти признаки выделены на рисунке 1.

Для классификации (кластеризации) текстовых документов разработано большое количество методов и алгоритмов.

Рассмотрим возможности и проблемы использования известных методов для анализа документов указанного типа.

Из представленных на рисунке 2 типов методов для анализа коротких неструктурированных документов при наличии динамического изменения характеристик рубрик в наибольшей степени подходят методы на основе нейросетей, вероятностные и генетические.

В то же время известные варианты данных методов имеют ограничения при автоматизированном анализе обращений граждан: нейросетевые методы достаточно сложны в обучении и связаны с построением большой трудномасштабируемой сети и сложным выбором метрики [6–8]; вероятностные классификаторы на коротких и неструктурированных текстовых документах на естественном языке не дают достаточную точность и эффективность [9–12]; генетические алгоритмы трудно обучаются и тяжело перестраиваются под динамические характеристики тезауруса [13].



В этой связи представляется целесообразным использование их модификаций.

Идея разработанного метода на основе весовых коэффициентов состоит в следующем:

- каждому слову, соответствующему предметной области, назначается вес;
- по умолчанию вес всех слов принимается равным;
- проводится обучение метода на некоторой первоначальной выборке документов, в ходе которого веса слов изменяются соответственно их значимости в контексте конкретной предметной области;
- проводится корректировка весовых коэффициентов экспертом и на начальном этапе, и в процессе работы системы, так как известна правильность результатов анализа.

Чтобы показать модифицированную статистическую часть метода, приведем пример стандартного вероятностного классификатора в виде формулы

$$J \leftarrow F(d_k) = \max_j \left(\sum_i \begin{cases} f_{i,j} \cdot |c_{ik}|, & \frac{|c_{ik}|}{|d_k|} \geq P_{i,j} \\ 0, & \frac{|c_{ik}|}{|d_k|} < P_{i,j} \end{cases} \right),$$

где $F(d_k)$ – максимальная сумма произведения частот употребления слов на количество их употребления в k -м документе, которая определяет предметную область J ; $f_{i,j}$ – частота употребления i -го слова в j -й предметной области; $|c_{ik}|$ – количество употреблений i -го слова в k -м текстовом документе; $|d_k|$ – количество слов в k -м документе; $P_{i,j}$ – порог употребления i -го слова в j -й предметной области.

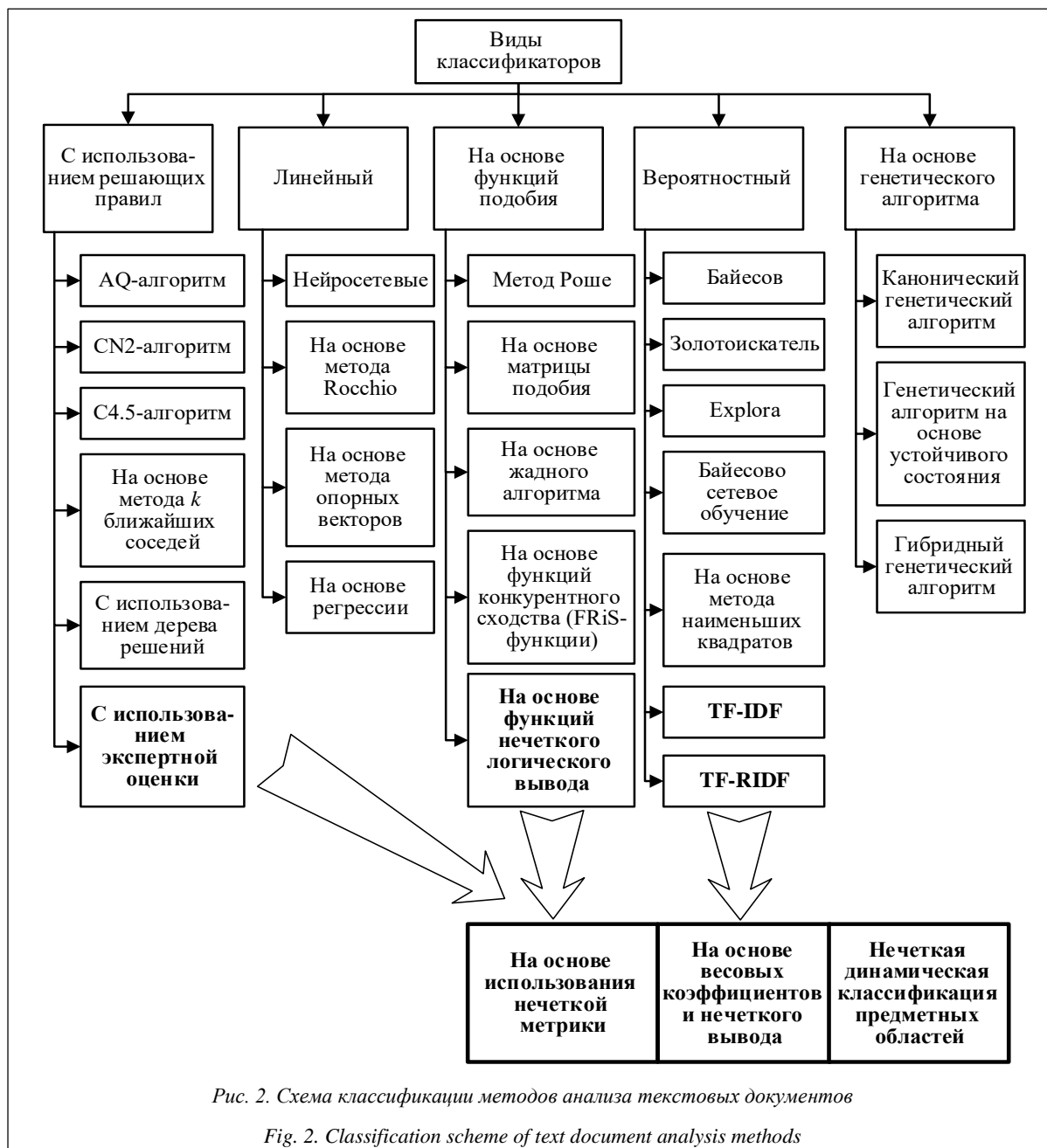
Математическую модель разработанного метода автоматической классификации можно представить в виде формулы

$$J \leftarrow W(d_k) = \max_j \sum_{i=1}^{n_k} K_{i,k} \cdot w_{i,j} \cdot |c_{i,k}|,$$

где $W(d_k)$ – максимальная сумма произведений весовых коэффициентов на количество употреблений в k -м документе (весовой коэффициент, в свою очередь, зависит еще от синтаксического коэффициента важности и актуальности информации); $w_{i,j}$ – весовой коэффициент i -го слова для j -й предметной области; $|c_{i,k}|$ – количество употреблений i -го слова в k -м документе; $K_{i,k}$ – синтаксический коэффициент важности i -го слова в k -м документе (определяется синтаксической значимостью слова в предложении).

В документах встречаются слова (общие), которые употребляются почти во всех предметных областях, они не несут информацию о предметной области документа. Следовательно, их веса необходимо сделать намного меньше других. Слова, встречающиеся только в одной предметной области (уникальные), являются самыми значимыми, и их веса будут значительно больше других. Еще остаются редкие слова, которые не являются ни уникальными, ни общими. Они несут некоторую информацию о предметной области, поэтому им назначаются промежуточные значения весовых коэффициентов. Алгоритм обучения проводит анализ БД ключевых слов и разбивает их на три категории: уникальные, редкие и общие. Далее эксперт выбирает нужное соотношение и значение весовых коэффициентов заданных трех типов ключевых слов. Для разграничения редких и общих слов вводится порог встречаемости: если слово встретилось в документах только одной предметной области, то это уникальные слова, если не во всех и меньше порога, то редкие, все остальные слова являются общими.

В ходе экспериментов были получены оптимальные значения весовых коэффициентов и порога отбора общеупотребительных слов: вес уникальных слов = 50, редких = 10, общих = 1, а порог отбора общих слов составляет 80 %. При данных характеристиках алгоритма обучения метод показывает наилучшие результаты.



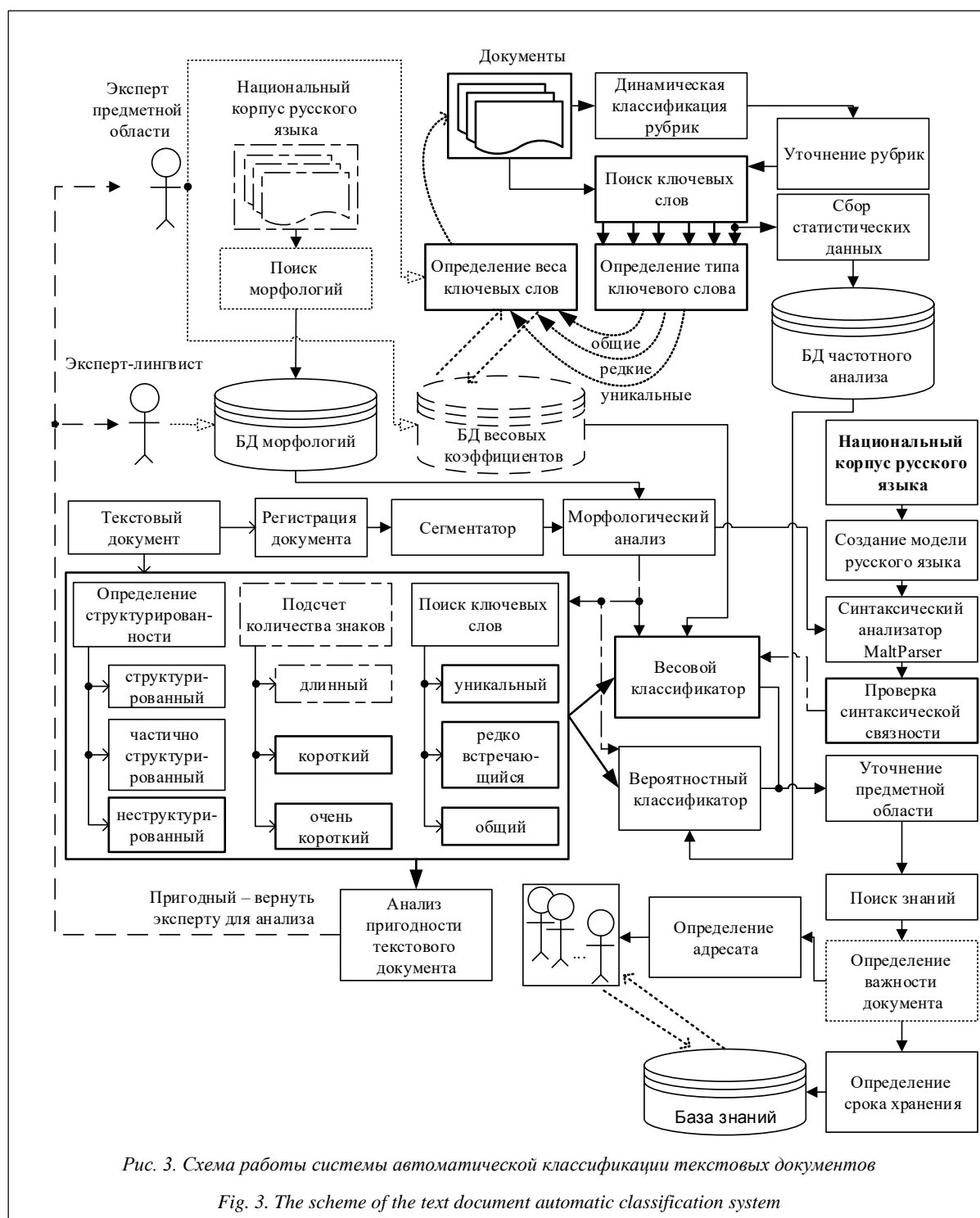
Можно выделить следующие преимущества весового алгоритма перед частотным алгоритмом:

- отсутствие порога частоты употребления слов позволяет распознавать короткие предложения, содержащие большое количество сокращений, цифр и минимум одно ключевое слово;
- использование весов, а не частот употребления увеличивает возможности для обучения алгоритма;
- при экспертной настройке базы знаний слов предметных областей удобнее работать с весами, чем с частотами и порогами употребления;
- возможность использовать слова, не принадлежащие напрямую к предметной области, но приписанные к ней с маленькими весовыми коэффициентами.

Недостатки весового алгоритма по сравнению с частотным алгоритмом:

- возрастает сложность процесса обучения;
- отсутствие порога распознавания предложения увеличивает вероятность распознавания документа, который относится к предметной области, неизвестной нашей системе;
- появление уникального слова с большим весом в другой предметной области сильно увеличит шанс на ошибочное распознавание документа.

Проблема динамического тезауруса в разработанном методе решается путем анализа сдвига кластеров предметных областей и своевременного запуска переобучения и подстройки метода. Методы динамической классификации изложены в [14–17].



Нечеткий логический вывод рассмотрен в статьях [18] и [19].

На рисунке 3 представлена схема функционирования системы автоматической классификации текстовых документов при использовании разработанного метода, предполагающего организацию обратной связи с экспертами.

Текстовый документ поступает на вход системы и проходит следующие этапы анализа: регистрация в системе, где получает уникальный номер

и озаглавляется необходимыми тэгами для дальнейшей работы; сегментация на слова, предложения, абзацы; морфологический анализ, определяющий лексические характеристики слов и морфемы; классификация документа по трем характеристикам и определение метода его анализа или обозначение документа как непригодного и возврат его экспертам для анализа (если документ короткий или очень короткий и неструктурированный, то используется разработанный метод анализа, в про-

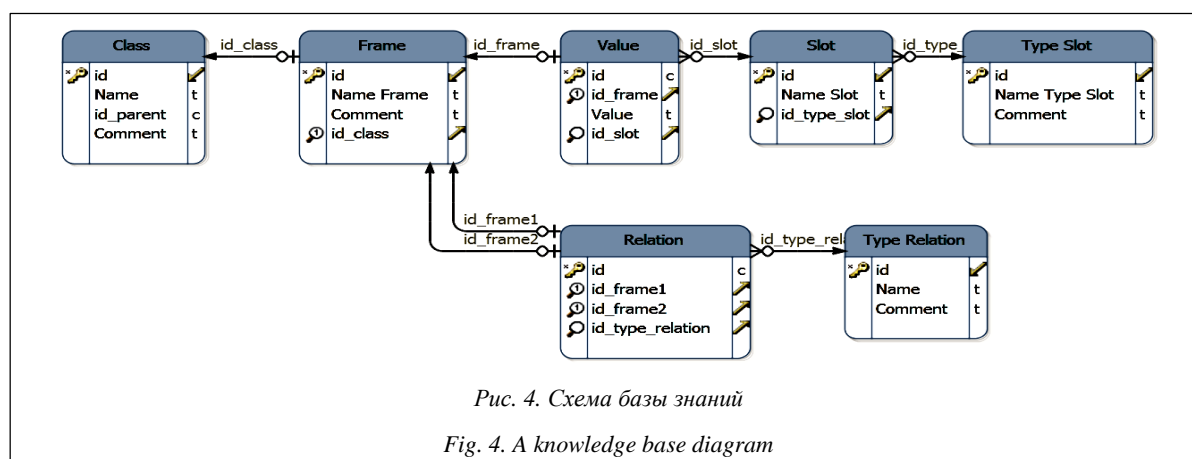


Рис. 4. Схема базы знаний

Fig. 4. A knowledge base diagram

тивном случае – вероятностный); уточнение предметной области; синтаксическая разметка документа (для разработанного метода) и проверка синтаксической связности; поиск знаний, определение важности документа и адресата, отправка документа ответственному лицу; сохранение полученных знаний в базу знаний для дальнейшего использования. База знаний реализуется в виде фрейм-овой реляционной БД, которая хранит извлеченные понятия и отношения между ними. На рисунке 4 представлена схема базы знаний.

Предметная область, к которой принадлежит извлеченный фрейм Frame, представлена в виде сущности Class. Relation описывает отношения между фреймами, а также задает тип отношений. Value предназначен для хранения характеристик фреймов и соотносится со слотами и их типами [20].

В качестве ПО для БД используется MS SQL Server 2008, функционала которого достаточно для реализации поставленных задач. Программирование остальных этапов анализа, за исключением MaltParser, осуществляется на Microsoft Visual C#.

Для морфологического анализа предварительно составляется морфологический словарь на Национальном корпусе русского языка.

Для поддержания актуальности тезауруса каждый раз происходит динамическое отслеживание рубрик и их уточнение.

Основываясь на недостатках разработанного и имеющихся методов, необходимо изменить весь процесс анализа текстовых документов, как предложено на рисунке 3. Весовой метод необходимо применять только при анализе коротких текстовых документов, следовательно, добавляется этап проверки длины текстового документа, который определяет метод анализа: один из известных методов или описанный выше.

Литература

1. Батура Т.В., Мурзин Ф.А., Проскуряков А.В. Программный комплекс для анализа данных из социальных сетей // Программные продукты и системы. 2015. № 4. С. 188–197.
2. Schutze H., Hull D.A., Pedersen J.O. A comparison of classifiers and document representations for the routing problem. Proc.

SIGIR-95, 18th ACM Inter. Conf., Seattle, USA, 1995, pp. 229–237.

3. Ng H.T., Goh W.B., Low K.L. Feature selection, perceptron learning, and a usability case study for text categorization. Proc. SIGIR-97, 20th ACM Intern. Conf. 1, Philadelphia, USA, 1997, pp. 67–73.

4. Dagan I., Karov Y., Roth D. Mistake-driven learning in text categorization. Proc. EMNLP-97, 2nd Conf., Providence, USA, 1997, pp. 55–63.

5. Joachims T. Text categorization with support vector machines: learning with many relevant features. Proc. ECML-98, 10th Europ. Conf., Chemnitz, Germany, 1998, pp. 137–142.

6. Lam S.L., Lee D.L. Feature reduction for neural network based text categorization. Proc. DASFAA-99, Taiwan, 1999, pp. 195–202.

7. Ruiz M., Srinivasan P. Hierarchical text categorization using neural networks. Information Retrieval, 2002, vol. 5, no. 1, pp. 87–118.

8. Yang Y., Liu X. A re-examination of text categorization methods. Proc. of SIGIR-99, 22nd ACM Inter. Conf., Berkeley, USA, 1999, pp. 42–49.

9. Козлов П.Ю. Сравнение частотного и весового алгоритмов автоматического анализа документов // Научное обозрение. 2015. № 14. С. 245–250.

10. Lewis D.D. Naive (Bayes) at forty: The independence assumption in information retrieval. Proc. ECML-98, 10th Europ. Conf., Chemnitz, Germany, 1998, p. 415.

11. Heckerman D. A tutorial on learning with bayesian networks. Learning in graphical models, 1999, pp. 301–354.

12. de Campos L.M., Romero A.E. Bayesian network models for hierarchical text classification from a thesaurus. Inter. Jour. of Approximate Reasoning, 2009, vol. 50, no. 7, pp. 932–944.

13. Wong M.L., Cheung K.S. Data mining using grammar based genetic programming and applications. Kluwer Acad. Publ., 2002, 228 p.

14. Гимаров В.А., Дли М.И., Круглов В.В. Задачи распознавания нестационарных образов // Изв. РАН. Теория и системы управления. 2004. № 3. С. 92–96.

15. Гимаров В.А., Дли М.И., Круглов В.В. Временная изменчивость образов // Вестн. МЭИ. 2003. № 2. С. 91–98.

16. Гимаров В.А., Дли М.И., Круглов В.В. Задачи динамической кластеризации // Системы управления и информационные технологии. 2005. Т. 18. № 1. С. 18–21.

17. Гимаров В.А., Дли М.И. Нейросетевой алгоритм классификации сложных объектов // Программные продукты и системы. 2004. № 4. С. 51–55.

18. Круглов В.В., Дли М.И., Голунов Р.Ю. Нечеткая логика и искусственные нейронные сети. М.: Наука. Физматлит, 2001. 224 с.

19. Круглов В.В., Дли М.И. Интеллектуальные информационные системы: компьютерная реализация логического вывода. М.: Физматлит, 2002. 256 с.

20. Хабаров С.П. Представление знаний с применением фреймов. URL: <http://www.habarov.spb.ru/bz07.htm> (дата обращения: 23.01.2017).

AUTOMATED ANALYSIS METHOD OF SHORT UNSTRUCTURED TEXT DOCUMENTS

P.Yu. Kozlov¹, Postgraduate Student, originaldod@gmail.com

¹ Smolensk Branch of the Moscow Power Engineering Institute, Energetichesky proezd 1, Smolensk, 214013, Russian Federation

Abstract. The paper considers the problem of an automated analysis of text documents in the executive and legislative authorities. It provides a characteristics group in order to classify text documents, their types, methods of analysis and rubricating. There is a list of the types of documents that need to be classified.

To analyze short unstructured text documents the authors propose to use a classification method based on weighting factors, expert information, fuzzy inference with a developed probabilistic mathematical model, a way of learning and experimentally chosen ratio of weight coefficients. The pre-developed method should be trained.

During learning the thesaurus words for each domain are divided into three types: unique, rare and common. The words are allocated with weights depending on the type. In order to maintain the relevance of weight and frequency coefficients it is proposed to use dynamic clustering. The developed method allows analyzing the disclosed documents, as well as taking into account thesaurus heading agility.

The paper presents a scheme of automatic classification system for unstructured text documents written in natural language. There might be various types of text documents: long, short, very short. Depending on the document type the system uses a corresponding method of analysis, which has the best indicators of accuracy and completeness of such text document analysis. MaltParser is a parser which is used here and trained on a national set of the Russian language. The result of the whole system work is a knowledge base, which includes all extracted knowledge and attitudes. The knowledge base is constantly updated and used by employees of the executive and legislative authorities to handle incoming requests.

Keywords: analysis automated analysis of texts, short texts unstructured, dynamic thesaurus.

References

1. Batura T.V., Murzin F.A., Proskuryakov A.V. A software package for a social network data analysis. *Programmnye produkty i sistemy* [Software & Systems]. 2015, no. 4, pp. 188–197 (in Russ.).
2. Schutze H., Hull D.A., Pedersen J.O. A comparison of classifiers and document representations for the routing problem. *Proc. 18th ACM Int. Conf. on Research and Development in Information Retrieval (SIGIR-95)*. Seattle, 1995, pp. 229–237.
3. Ng H.T., Goh W.B., Low K.L. Feature selection, perceptron learning, and a usability case study for text categorization. *Proc. 20th ACM Int. Conf. on Research and Development in Information Retrieval (SIGIR-97)*. Philadelphia, 1997, pp. 67–73.
4. Dagan L., Karov Y., Roth D. Mistake-driven learning in text categorization. *Proc. 2nd Conf. on Empirical Methods in Natural Language Processing (EMNLP-97)*. Providence, RI, 1997, pp. 55–63.
5. Joachims T. Text categorization with support vector machines: learning with many relevant features. *Proc. 10th European Conf. on Machine Learning (ECML-98)*. Chemnitz, Germany, 1998, pp. 137–142.
6. Lam S.L., Lee D.L. Feature reduction for neural network based text categorization. *Proc. DASFAA-99*. Taiwan, 1999, pp. 195–202.
7. Ruiz M., Srinivasan P. Hierarchical Text Categorization Using Neural Networks. *Information Retrieval*. 2002. vol. 5, no. 1, pp. 87–118.
8. Yang Y., Liu X. A re-examination of text categorization methods. *Proc. 22nd ACM Int. Conf. on Research and Development in Information Retrieval (SIGIR-99)*. Berkeley, CA, 1999, pp. 42–49.
9. Kozlov P.Yu. Comparing frequency and weight algorithms for automatic document analysis. *Nauchnoe obozrenie* [Scientific Review]. 2015, no. 14, pp. 245–250 (in Russ.).
10. Lewis D.D. Naive (Bayes) at forty: The independence assumption in information retrieval. *Proc. 10th European Conf. on Machine Learning (ECML-98)*. Chemnitz, Germany, 1998, p. 415.
11. Heckerman D. A Tutorial on Learning With Bayesian Networks. *Learning in graphical models*. 1999, pp. 301–354.
12. de Campos L.M., Romero A.E. Bayesian network models for hierarchical text classification from a thesaurus. *Int. Jour. of Approximate Reasoning*. 2009, vol. 50, no. 7, pp. 932–944.
13. Wong M.L., Cheung K.S. *Data Mining Using Grammar Based Genetic Programming and Applications*. Kluwer Academic Publ., 2002, 228 p.
14. Gimarov V.A., Dli M.I., Kruglov V.V. Unstable image recognition problems. *Izvestiya Ross. akademii nauk. Teoriya i sistemy upravleniya* [Jour. of Computer and Systems Sciences International]. 2004, no. 3, pp. 92–96 (in Russ.).
15. Gimarov V.A., Dli M.I., Kruglov V.V. Temporary image variation. *Vestnik MEI* [MPEI Vestnik]. 2003, no. 2, pp. 91–98 (in Russ.).
16. Gimarov V.A., Dli M.I., Kruglov V.V. Dinamic clustering problems. *Sistemy upravleniya i informatsionnye tekhnologii* [Control Systems and IT]. 2005, vol. 18, no. 1, pp. 18–21 (in Russ.).
17. Gimarov V.A., Dli M.I. A neural network algorithm for classification of complex text objects. *Programmnye produkty i sistemy* [Software & Systems]. 2004, no. 4, pp. 5–56 (in Russ.).
18. Kruglov V.V., Dli M.I., Golunov R.Yu. *Nechetkaya logika i iskusstvennyye neyronnyye seti* [Fuzzy logic and artificial networks]. Moscow, Nauka, Fizmatlit Publ., 2001.
19. Kruglov V.V., Dli M.I. *Intellektualnye informatsionnye sistemy: kompyuternaya realizatsiya logicheskogo vyvoda* [Intelligent Systems: Computer Implementation of a Logical Inference]. Moscow, Fizmatlit Publ., 2002.
20. Khabarov S.P. *Predstavlenie znaniy s primeneniem freymov* [Knowledge Representation Using Frames]. Available at: <http://www.habarov.spb.ru/bz/bz07.htm> (accessed January 29, 2017).

УДК 519.688
DOI: 10.15827/0236-235X.117.106-111

Дата подачи статьи: 01.08.16
2017. Т. 30. № 1. С. 106–111

ИСПОЛЬЗОВАНИЕ ГЕОМЕТРИИ СЦЕНЫ ДЛЯ УВЕЛИЧЕНИЯ ТОЧНОСТИ ДЕТЕКТОРОВ

*Е.В. Шальнов, аспирант, eshalnov@graphics.cs.msu.ru
(Московский государственный университет им. М.В. Ломоносова,
Ленинские горы, 1-52, г. Москва, 119991, Россия);*

*А.С. Конушин, к.ф.-м.н., доцент, ktosh@graphics.cs.msu.ru
(Московский государственный университет им. М.В. Ломоносова,
Ленинские горы, 1-52, г. Москва, 119991, Россия;
Национальный исследовательский университет «Высшая школа экономики»,
ул. Мясницкая, 20, г. Москва, 101000, Россия)*

Ключевым элементом любых систем интеллектуальной видеоаналитики является алгоритм выделения, или детектирования объектов в видео. Недостаточно высокая скорость и точность существующих алгоритмов детектирования являются существенными сдерживающими факторами распространения технологий видеоаналитики.

В данной работе предлагается новый алгоритм повышения скорости и точности работы детекторов, основанных на подходе скользящего окна, за счет учета геометрических свойств сцены. В зависимости от расположения камеры относительно сцены для каждой области изображения можно определить, какого размера может быть изображение искомого объекта в данной области. Окна других размеров не могут соответствовать искомым объектам в сцене, поэтому их можно пропускать и за счет этого увеличивать скорость работы детектора. Предлагаемый алгоритм позволяет определить допустимые размеры объекта для каждой области изображения. Сутью алгоритма является нейронная сеть, которая для таких заданных параметров, как калибровка камеры, размер и положение объекта на снимке, определяет, правдоподобна ли данная сцена. Нейронная сеть обучается на множестве синтетических сцен, что позволяет ей работать для произвольных камер. С помощью нейронной сети для каждой видеопоследовательности строится карта допустимых размеров объектов. Детектор затем применяется только к допустимым фрагментам, которые составляют часть от всего множества фрагментов.

Экспериментальная оценка предложенного алгоритма на реальных данных показала, что он позволяет повысить скорость работы детектора на 70 % при одновременном увеличении точности его работы.

Ключевые слова: нейронные сети, компьютерное зрение, обнаружение объектов, видеоаналитика, компьютерная графика, распознавание образов.

Обнаружение и локализация объектов интереса в видео – классическая задача компьютерного зрения. Несмотря на долгую историю развития методов обнаружения [1–3], ключевой подход к решению, называемый методом скользящего окна и заключающийся в применении классификатора к разным частям (окнам) изображения, остается неизменным. Классификатор определяет, содержит ли рассматриваемая область изображения объект интереса или нет. Обычно рассматривают окна, отличающиеся не только положением, но и размерами.

Количество таких окон может быть очень велико. Например, при поиске голов людей на изображении fullHD разрешения (1920×1080) (рис. 1) алгоритм [2] классифицирует 1 310 859 окон. Их обработка требует значительных вычислительных ресурсов. Эта проблема существенно ограничивает широкое применение алгоритмов видеоаналитики на практике.

В то же время можно заметить, что в большинстве анализируемых окон размеры человека в сцене неправдоподобны. Например, маловероятно в сцене, представленной на рисунке 1, увидеть голову человека, занимающую половину изображения. Такие априорные знания могут существенно сократить количество обрабатываемых окон и увеличить скорость обработки видеоданных. Для этого необходимо разработать алгоритм классифи-

кации окон изображений на ложные и правдоподобные в зависимости от положения камеры. К сожалению, в настоящий момент для произвольной сцены такого алгоритма фильтрации обнаружений не существует.

В данной работе представлен алгоритм определения ложных срабатываний детектора головы человека, использующий информацию о наблюдаемой сцене. Авторы ограничиваются рассмотрением только простых сцен:

- камера снимает плоскую сцену, то есть наблюдаемая поверхность земли является горизонтальной плоскостью;
- люди могут находиться только на поверхности земли;
- дисторсией камеры можно пренебречь.

Эти предположения существенно ограничивают набор рассматриваемых сцен. Так, сцена, содержащая лестницу, не подходит под описанные ограничения. В то же время рассматриваемый класс сцен является наиболее распространенным в видеонаблюдении.

Основной вклад данной работы:

- предложен метод построения классификатора обнаружений на невозможные с геометрической точки зрения и правдоподобные;
- показан способ построения синтетической выборки для обучения классификатора голов человека;



Рис. 1. Результаты работы детектора голов людей на изображении (красным цветом обозначены обнаружения, классифицированные предложенным методом как ложные)

Fig. 1. An image of human head detector results. False detections are shown in red

- показано, что построенный алгоритм классификации обнаружений позволяет повысить точность обнаружения голов людей;
- предложен способ интеграции классификатора обнаружений с используемым детектором голов, который позволил повысить скорость и точность обработки данных.

Смежные работы

Задача построения детектора объектов на изображении всегда интересовала исследователей в области компьютерного зрения. Обычно на разрабатываемые алгоритмы накладывались требования по времени работы и количеству ложных срабатываний. Эти требования зачастую противоречили друг другу. Действительно, часто повышение точности классификатора окон приводит к повышению его вычислительной сложности. Для практического применения в видеонаблюдении скорость обработки данных является ключевым параметром, поэтому многие разработчики стремились найти способы понижения вычислительной сложности детекторов при сохранении их качества. Можно выделить два основных направления работы в этой области: построение быстрого классификатора и уменьшение количества рассматриваемых окон.

Первые работы по ускорению детектирования посвящены ускорению применяемого классификатора. Авторы [1] предложили использовать каскад простых классификаторов для детектирования лиц на изображениях. Первые этапы каскада отбрасы-

вают большое количество простых окон, не содержащих лиц. Предложенная идея оказалась настолько эффективной, что каскадные детекторы стали применяться даже в цифровых фотоаппаратах. Одним из важных недостатков такого подхода является отсутствие возможности изменять соотношение точность/полнота для уже построенного классификатора. Авторы работы [4] преодолели это ограничение, изменив структуру каскада. Они разделили построение простых классификаторов на каждом этапе каскада и выбор границы для разделения положительных и отрицательных примеров. В [5] ускорение классификатора достигнуто за счет вычисления признаков лишь на разреженной пирамиде изображений. На промежуточных слоях предложено восстанавливать признаки с помощью интерполяции. В данной работе предложен алгоритм понижения вычислительной сложности детектора, который не зависит от типа используемого классификатора окон, поэтому его можно использовать совместно с быстрыми классификаторами.

Другое направление по ускорению обнаружения на изображениях – уменьшение количества рассматриваемых окон. Авторы работы [6] используют корреляцию откликов классификатора в соседних окнах для выделения регионов изображения, где могут находиться объекты. Для этого разреженное множество окон классифицируют на первых этапах обработки. В связи с существенным развитием нейросетевых алгоритмов классификации изображений [7–10] сверточные нейронные сети стали применять и для задачи обнаружения

объектов на изображении. Обычно нейросетевые классификаторы требуют больших вычислительных ресурсов, поэтому в работе [11] было предложено классифицировать лишь небольшое подмножество выбранных окон. В работах [3, 12] авторы развили предыдущую идею и предложили разбить классификатор на этапы классификации и уточнения положения объекта. Это позволило увеличить размер окон и уменьшить их количество.

Предложенный в данной работе алгоритм может быть интегрирован с любым из методов уменьшения количества обрабатываемых окон, поскольку дает априорную оценку областей, где могут находиться объекты интереса.

Предложенный метод

В данной работе создан алгоритм классификации обнаружений детектора голов людей на корректные и невозможные с геометрической точки зрения. Построенный алгоритм опирается на информацию о положении и параметрах камеры. Важно отметить, что ключевым требованием к разработанному классификатору является инвариантность к классификатору окон, используемому детектором. Это позволяет построить алгоритм фильтрации обнаружений для любого детектора. Формально входом алгоритма являются положение камеры и признаки классифицируемого обнаружения. Под обнаружением понимается прямоугольник, ограничивающий изображение объекта, найденного детектором.

Предложенный метод предусматривает три этапа построения классификатора результатов детектора:

- 1) построение синтетической выборки изображений;
- 2) построение признаков обнаружений, инвариантных для синтетических и реальных данных;
- 3) обучение классификатора.

Рассмотрим подробнее каждый из этих этапов для построения классификатора обнаружений детектора голов людей.

Построение обучающей выборки

Для обучения классификатора обнаружений необходима обучающая выборка. Авторы не могли использовать размеченные данные видеонаблюдения из-за сложности их сбора, поэтому с помощью компьютерной графики построили синтетическую выборку, моделирующую сценарий видеонаблюдения. Выборка состоит из 30 179 сцен, все они определяются положением камеры. Каждая сцена состоит из плоскости земли, стоящих на ней людей и камеры. Использовалась упрощенная модель камеры, которая определяется углами наклона и поворота камеры, фокусным расстоянием и высотой камеры над плоскостью земли. Для построения

каждой сцены параметры камеры выбирались случайным образом из равномерного распределения (см. таблицу).

Параметры камеры

Camera parameters

Обозначение	Параметр	Диапазон значений
h	Высота	$[0; 20]$
P	Наклон	$\left[\frac{\pi}{2}; \frac{11\pi}{12} \right]$
R	Поворот	$\left[-\frac{\pi}{12}; \frac{\pi}{12} \right]$
F	Фокусное расстояние	$[0; 5000]$

Для моделирования людей в сцене использовались результаты работы [13]. Каждая построенная сцена содержит не менее 200 человек, размещенных в случайных положениях плоскости земли.

Извлечение признаков обнаружений

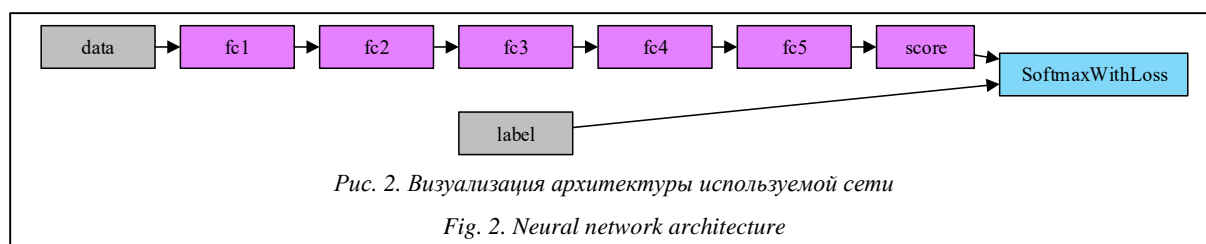
Обучение алгоритмов машинного обучения только на синтетической выборке может привести к эффекту переобучения, поэтому важным этапом обработки является построение признаков, инвариантных к типу обрабатываемых данных. В качестве таких признаков выбраны параметры ограничивающего прямоугольника, обнаруженного детектором. Для задачи фильтрации обнаружений головы была использована оптимизированная версия детектора голов людей [2] из-за высокой скорости его работы.

Модель [13] позволила исключить ложные обнаружения из обучающей выборки за счет информации о положении ключевых точек на человеке.

Построение классификатора

Формально задача классификации обнаружений ложных срабатываний детектора по построенной выборке относится к классу задач поиска аномалий. Отличительной особенностью таких задач является отсутствие примеров отрицательного класса (ошибки детектора). Сведем данную задачу к задаче классификации, предложив простую модель распределения ошибочных результатов работы детектора, состоящую из смеси двух распределений.

Первое моделирует случайные ошибки детектора. Для этого используется равномерное распределение обнаружений (их положений и размеров) по изображению. В качестве второго компонента смеси взято распределение обнаружений в построенной выборке. Это позволяет алгоритму отличить обнаружения, характерные для сцен. Эксперименты показали, что наилучшие результаты классификации получаются при смешивании этих распределений в отношении 1:9.



Для построения классификатора обнаружений детектора использована нейронная сеть, представленная на рисунке 2. Она состоит из полносвязных слоев, после которых используется нелинейная функция ReLu.

Тестирование и оценка

Предложенная сеть обучена на синтетической выборке, состоящей из 24 143 сцен, с помощью библиотеки caffe [14]. Скорость обучения понижалась каждые 1 500 итераций с параметром γ , равным 0,95.

Проведены оценка качества предложенного алгоритма на синтетических и реальных данных, а также интеграция его с используемым детектором.

Тестирование на синтетических данных. Тестирование проведено на синтетической выборке. Тестовая выборка состояла из 6 036 различных сцен, не использованных при обучении. На рисунке 3а показано качество классификации обнаружений на синтетической тестовой выборке.

Тестирование на реальных данных. Для проведения тестирования необходимо знать положение камеры для каждого кадра выборки. Это затрудняет использование стандартных баз размеченных изображений.

Использована выборка TownCentre [15], так как она содержит большую часть необходимой информации. К сожалению, параметры положения камеры, представленные в выборке, оказались неправдоподобными. Это может быть связано с использованием единицы измерения в мировой системе координат, отличной от метра. В данной сцене высота камеры над землей приблизительно равна 8 метрам.

Авторы применили фильтрацию обнаружений детектора с порогом 0,25. Обнаружение считалось правильным, если его пересечение с регионом головы человека в разметке занимало не менее 25 % их объединения. Сравнение качества исходного и полученного детекторов (рис. 3б) выявило, что предложенный алгоритм фильтрации позволяет увеличить точность без существенного уменьшения полноты обнаружения.

Интеграция с детектором

Использовался построенный фильтр обнаружений детектора для ускорения его работы. Действительно, можно оценить, какую область изображе-

ния необходимо обрабатывать на каждом кадре.

Построены маски областей, которые необходимо обрабатывать на каждом уровне пирамиды изображений для камеры, соответствующей выборке TownCentre (рис. 4, первая строка). Результаты показывают, что в сцене необходимо обработать только небольшое подмножество окон. Например, для сцены TownCentre достаточно обработать лишь 21,44 % всех окон.

Базовый детектор расширен возможностью обрабатывать только те строки изображения на каждом уровне пирамиды, где правдоподобно обнаружение головы человека. Пример обрабатываемых областей представлен на рисунке 4 (вторая строка). Это соответствует обработке 24,03 % всех окон.

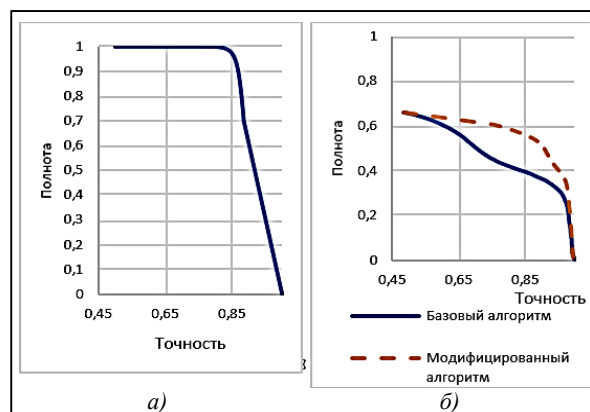


Рис. 3. Качество фильтрации обнаружений на синтетической тестовой выборке (а) и качество работы детектора до и после использования фильтра обнаружений (б)

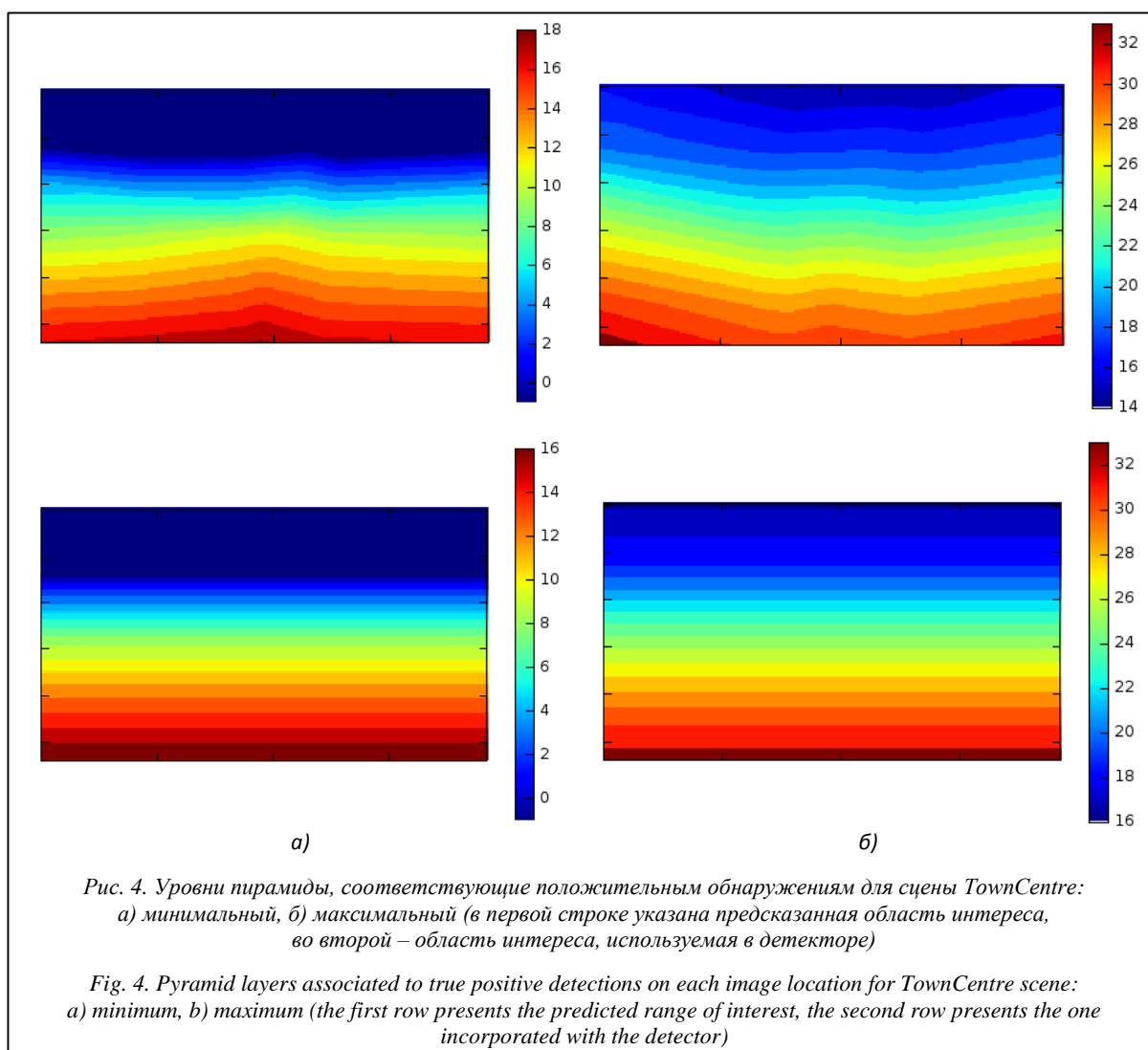
Fig. 3. Filtration quality on the synthetic test set (a) and detector performance quality with and without filtration (b)

Выделение областей интереса является эффективным способом повышения производительности детектора. Обработка области интереса на сцене TownCentre позволила повысить скорость обработки данных с 20,03 до 34,36 кадра в секунду.

Полученный результат наиболее важен для систем видеонаблюдения, где параметры камеры резко изменяются и могут быть оценены один раз.

Заключение

В данной работе представлен эффективный способ фильтрации обнаружений детектора, осно-



ванный на использовании информации о положении камеры.

Предложенный метод позволил повысить точность и скорость работы детектора.

Работа выполнена при поддержке РФФИ, грант № 15-31-20596, и Сколковского института науки и технологий, договор № 081-R приложение А2.

Литература

1. Viola P., Jones M. IEEE. Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition, 2001, CVPR 2001, Proc. IEEE Comp. Society Conf. 2001, vol. 1, p. 511.
2. Prisacariu V., Reid I. FastHOG-a real-time GPU implementation of HOG. Department of Eng. Sc., 2009, vol. 2310, no. 9, pp. 325–332.
3. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems, 2015, pp. 91–99.
4. Bourdev L., Brandt J. IEEE. Robust object detection via soft cascade. Proc. 2005 IEEE Computer Society Conf. CVPR'05, 2005, vol. 2, pp. 236–243.
5. Dollár P., Belongie S., Perona P. The fastest pedestrian detector in the west. BMVC, 2010, vol. 2, p. 7.
6. Dollár P., Appel R., Kienle W. Crosstalk cascades for frame-rate pedestrian detection. Springer, 2012, pp. 645–659.
7. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
8. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
9. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385, 2015.
10. Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the inception architecture for computer vision. arXiv preprint arXiv:1512.00567, 2015.
11. Girshick R., Donahue J., Darrell T., Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. Proc. IEEE Conf. on Comp. Vision and Pattern Recognition, 2014, pp. 580–587.
12. Girshick R. Fast r-cnn. Proc. IEEE Intern. Conf. on Comp. Vision, 2015, pp. 1440–1448.
13. Pishchulin L., Wuhler S., Helten T., Theobalt C., Schiele B. Building statistical shape spaces for 3d human modeling. arXiv preprint arXiv:1503.05860, 2015.
14. Jia Y., Shelhamer E., Donahue J., Karayev S., Long J., Girshick R., Guadarrama S., Darrell T. ACM. Caffe: Convolutional architecture for fast feature embedding. Proc. 22nd ACM Intern. Conf. on Multimedia, 2014, pp. 675–678.
15. Benfold B., Reid I., IEEE. Stable multi-target tracking in real-time surveillance video. Proc. IEEE Conf. CVPR, 2011, pp. 3457–3464.

SCENE GEOMETRY FOR DETECTOR PRECISION IMPROVEMENT

E.V. Shalnov¹, Postgraduate Student, eshalnov@graphics.cs.msu.ru

A.S. Konushin^{1,2}, Ph.D (Physics and Mathematics), Associate Professor, ktosh@graphics.cs.msu.ru

¹ Lomonosov Moscow State University, Leninskie Gory, Moscow, 119991, Russian Federation

² National Research University, Higher School of Economics,
Myasnitskaya St., 20, Moscow, 101000, Russian Federation

Abstract. Object detection algorithms are the key component of any intelligent video content analysis systems. High computation requirements and low precision of existing methods restrain widespread acceptance of intelligent video content analysis.

The paper introduces a novel algorithm that accelerates existing sliding window object detectors and increases their precision. This approach is based on the geometric properties of an observed scene. If the camera position in the scene is known, we can determine feasible sizes of detected objects in each location of an input image. Windows of other sizes cannot correspond to objects in a scene and thus could be skipped. It significantly decreases computation time. The proposed algorithm estimates feasible sizes of object for each location of an input image. We apply Neural Network (NN) to solve this task. A NN takes camera calibration parameters and window parameters as the input and determines if this configuration feasible or not. We train the NN on the synthetic dataset. It allows us to take into account a huge range of camera calibration parameters. We apply the NN to construct a map of feasible object sizes for the input scene.

Thus the detector processes the feasible subset of windows. The performed evaluation reveals that the proposed algorithm accelerates processing by 70 % and increases precision of a detector.

Keywords: neural networks, computer vision, object detection, video content analysis, computer graphics, pattern recognition.

Acknowledgements. The research has been supported by RFBR, grant no. 15-31-20596 and Skolkovo Institute of Science and Technology, the contract no. 081-R, Annex A2.

References

1. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features. *Proc. of the 2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2001)*. 2001, vol. 1, p. 511.
2. Prisacariu V., Reid I. *FastHOG-a real-time GPU implementation of HOG*. Department of Eng. Sc., 2009, vol. 2310, no. 9, pp. 325–332.
3. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*. 2015, pp. 91–99.
4. Bourdev L., Brandt J. Robust object detection via soft cascade. *Proc. 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR '05)*. 2005, vol. 2, pp. 236–243.
5. Dollár P., Belongie S., Perona P. The Fastest Pedestrian Detector in the West. *BMVC*. 2010, vol. 2, p. 7.
6. Dollár P., Appel R., Kienzle W. *Crosstalk cascades for frame-rate pedestrian detection*. Springer Publ., 2012, pp. 645–659.
7. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*. 2012, pp. 1097–1105.
8. Simonyan K., Zisserman A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv preprint arXiv:1409.1556, 2014.
9. He K., Zhang X., Ren S., Sun J. *Deep Residual Learning for Image Recognition*. arXiv preprint arXiv:1512.03385, 2015.
10. Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. *Rethinking the Inception Architecture for Computer Vision*. arXiv preprint arXiv:1512.00567, 2015.
11. Girshick R., Donahue J., Darrell T., Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*. 2014, pp. 580–587.
12. Girshick R. Fast r-cnn. *Proc. IEEE Int. Conf. on Computer Vision*. 2015, pp. 1440–1448.
13. Pishchulin L., Wuhler S., Helten T., Theobalt C., Schiele B. *Building Statistical Shape Spaces for 3d Human Modeling*. arXiv preprint arXiv:1503.05860, 2015.
14. Jia Y., Shelhamer E., Donahue J., Karayev S., Long J., Girshick R., Guadarrama S., Darrell T. Caffe: Convolutional architecture for fast feature embedding. *Proc. 22nd ACM Int. Conf. on Multimedia*. 2014, pp. 675–678.
15. Benfold B., Reid I. Stable multi-target tracking in real-time surveillance video. *Computer Vision and Pattern Recognition (CVPR 2011)*, IEEE Conf., 2011, pp. 3457–3464.

УДК 331.446.3

DOI: 10.15827/0236-235X.117.112-118

Дата подачи статьи: 11.07.16

2017. Т. 30. № 1. С. 112–118

ПРОГРАММНЫЙ АГЕНТ ОПРЕДЕЛЕНИЯ ПСИХОЛОГИЧЕСКОГО СОСТОЯНИЯ ОБУЧАЕМОГО В СИСТЕМАХ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

Е.А. Хрянин, аспирант, главный инженер, evgeshah@list.ru

(Вологодский государственный университет, ул. Ленина, 15, г. Вологда, 160000, Россия;

Банк «Воложжанин», ул. Батюшкова, 11, г. Вологда, 160000, Россия);

А.Н. Швецов, д.т.н., профессор, smithv@mail.ru

(Вологодский государственный университет, ул. Ленина, 15, г. Вологда, 160000, Россия)

Рассматривается проблема применения программных агентов для оценки психологического состояния студентов в системе дистанционного обучения. Гипотеза исследования: чем лучше психологически материал подходит обучаемому, тем быстрее и качественнее он будет усвоен. Требуется разработать автоматический алгоритм подбора материала. Описывается разработанная система дистанционного обучения, создаваемая более 5 лет и апробированная в одном из государственных вузов.

Кратко описана реализация системы дистанционного обучения: схема взаимодействия агентов, основные таблицы БД, реализация серверной и пользовательской частей. Описываются метод и алгоритм определения перцептивной модальности обучаемого в ходе психологического тестирования. Используются статистические методы для предсказания вероятности входа в систему обучаемым (на основе данных статистики). Предложены весовые коэффициенты частоты использования системы дистанционного обучения обучаемыми для принятия решений агентом определения психологического состояния.

Создан алгоритм автоматического решения о необходимости тестирования. Проведено исследование на основе трех групп с участием более 90 человек: контрольная группа, группа с рекомендацией в выборе материала и группа, для которой агент сам выбирает материал. Выведены формулы расчета перцептивной модальности для нескольких последовательных измерений. Приведен пример уточнения расчета при получении противоречивых данных. Эксперимент показал положительные результаты при работе в рекомендательном режиме. С контрольным тестом справились более 61 % обучаемых, а усложненную задачу решили более половины группы (около 42 % и 12 % в контрольной группе соответственно).

Сделан вывод о целесообразности применения агента определения психологического состояния в системах дистанционного обучения.

Ключевые слова: *агентно-ориентированный подход, интеллектуальная система, система дистанционного обучения, диагностика психологического состояния, перцептивная модальность, оценка результатов обучения, РНР, MySQL, MVC.*

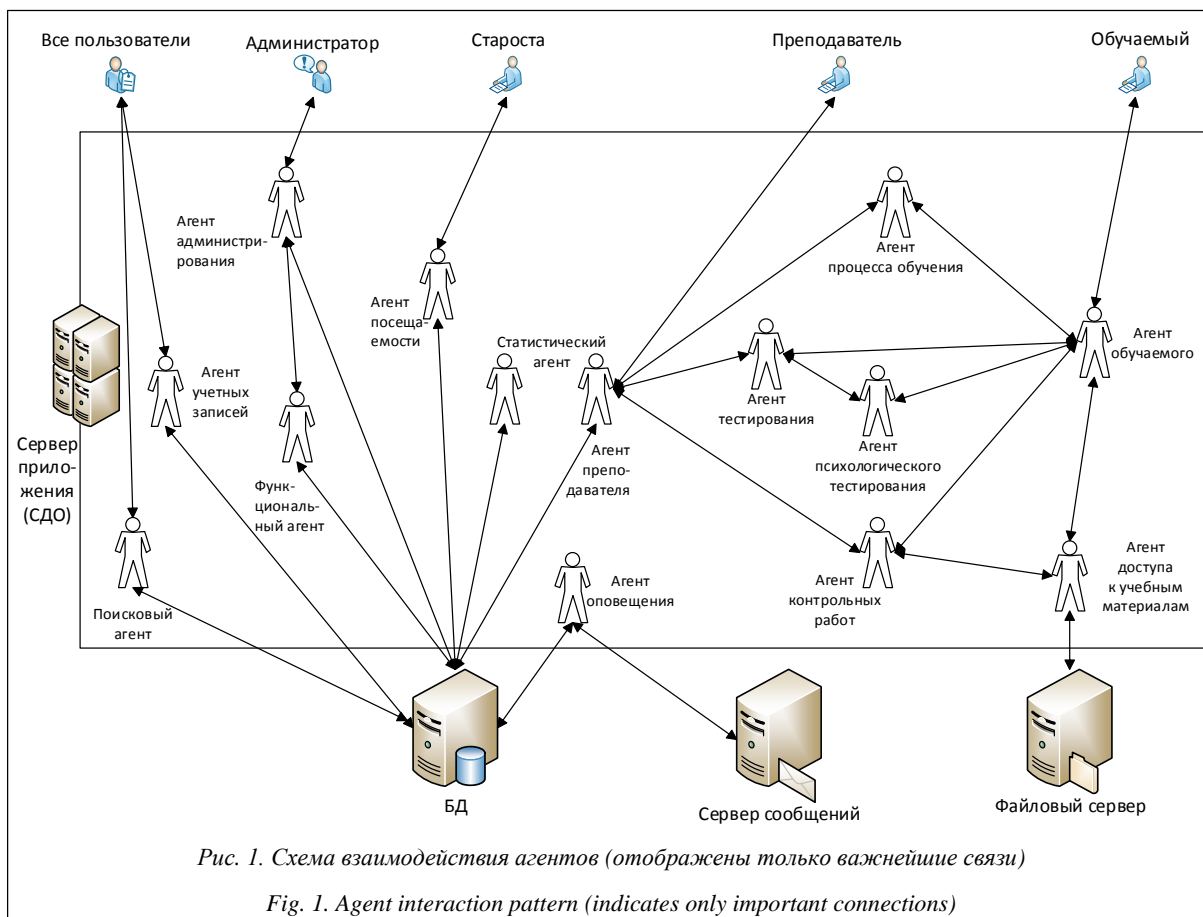
Объем знаний, необходимых для успешного обучения и последующего карьерного роста, увеличивается с каждым годом. Сложность и количество изучаемого материала растут, при этом сроки обучения остаются неизменными или сокращаются. Обычной практикой в вузе стало использование *систем дистанционного обучения* (СДО) [1]. Последние десятилетия активно используются СДО, построенные на агентно-ориентированном подходе [2–4]. Для обеспечения индивидуального подхода в предоставлении материалов необходим агент определения психологического состояния обучаемого в СДО.

В связи с этим целесообразно исследовать возможности агента, определяющего психологическое состояние обучаемого и предлагающего адаптированный материал для обучения. Гипотеза состоит в следующем: если дать обучаемому материал, который психологически лучше ему подходит, то обучение будет происходить быстрее и с лучшим усвоением, в результате на каждую тему обучения будет тратиться меньше времени. Вопросы оценки состояния рассматривались в других работах [5, 6], но применение такого агента к СДО не исследовано.

Объект работы – многофункциональная дистанционная агентно-ориентированная система обучения. Рассматриваемый агент является частью данной СДО [7]. Система позволяет вести базу образовательного материала, публиковать и принимать контрольные работы, проводить тестирование в рамках лекций, тем и дисциплин. Дополнительными возможностями системы являются учет посещаемости студентов на основе данных от нескольких независимых источников и отслеживание результатов их обучения. В систему интегрированы функции оповещения обучаемых и преподавателей о событиях, происходящих в ней.

Программная реализация СДО

Система представляет собой совокупность агентов различной направленности. Агенты анализируют собственное предметное поле, общаются между собой и дают пользователю то, что ему требуется в данный момент времени [8]. Комплекс состоит из следующих агентов: агент администрирования, агент учетных записей, агент обучаемого, агент преподавателя, агент процесса обучения, агент доступа к учебным материалам, агент кон-



трольных работ, агент тестирования, агент посещения, агент оповещения, статистический агент, агент психологического анализа, функциональный агент и поисковый агент (рис. 1).

Система была апробирована в одном из государственных вузов, где использовалась при очном и заочном обучении.

СДО написана на языке PHP без использования различных фреймворков для ускорения работы. Благодаря этому система моментально генерирует HTML-страницы на скромном оборудовании: ОС CentOS 6, процессор уровня Celeron 2000 МГц, 512 МБ ОЗУ, 30 МБ HDD для установки СДО.

В качестве основной СУБД используется MySQL, обеспечивающая достаточную производительность для решаемых задач. Тем не менее есть возможность быстро конвертировать ее для работы с СУБД PostgreSQL. Схема с основными таблицами в БД размещена по ссылке http://www.swsys.ru/uploaded/image/2017_1/2017-1-dop/1.jpg.

Основная таблица – Пользователи, так как весь контент сосредоточен вокруг них. Пользователь может быть администратором, преподавателем, студентом, старостой, при этом уровни доступа можно комбинировать. В зависимости от уровня доступа пользователь располагает теми или иными функциями и возможностями. Сама БД является нормализованной, что исключает избыточность данных и потенциально исключает ошибки. Право

записи в каждую из таблиц закреплено за определенными агентами. Например, агент преподавателя не может создавать пользователей, однако имеет возможность добавлять записи в таблицу дисциплин и изменять их. Для быстрой работы агентов требуются временные таблицы, которые они генерируют во время своей работы. Это обеспечивает обработку данных агентами независимо друг от друга, а также защиту основной БД от некорректных записей.

В качестве фронтенда используются HTML5, CSS3 и jQuery. Для реализации пользовательского интерфейса были применены фреймворк Bootstrap с некоторыми изменениями, а также интеграция с Google Docs (http://www.swsys.ru/uploaded/image/2017_1/2017-1-dop/2.jpg).

Система является кроссплатформенной как для серверной части, так и для клиентской, а поддержка современных технологий позволяет комфортно ею пользоваться и на стационарных компьютерах, и на смартфонах. СДО изначально создавалась с использованием идеологии MVC, поэтому имеет модульную архитектуру и возможность развиваться без изменения существующей кодовой базы.

За три года работы системы размер основной БД составил около 50 МБ: более 1 500 пользователей, более 500 дисциплин, 1 000 лекций, 750 контрольных работ, 1 000 вопросов для тестов. Размер

файлового хранилища для лекций, контрольных работ и их решений – более 4 Гб при почти 19 тыс. загруженных файлов. Среднегодовое количество обучаемых – около 800 человек.

Исследование. Исследование проводилось на базе СДО. Рассматривалось влияние на успеваемость рекомендаций от агента определения психологического состояния обучаемого. Изучалось, насколько эффективно агент определяет тип восприятия обучаемого и может предложить для него подходящий материал. В СДО могут применяться следующие обучающие материалы: видеолекции, презентации, разбор примеров, аудиолекции, общение с преподавателем, текст, тест, обучение в виде игры, представления в виде слайдов, графиков и формул [9].

Для определения перцептивной модальности обучаемых использовались несколько методик, адаптированных для СДО. За основу взяты методики Н.Л. Васильевой [10] и С. Ефремовой [11].

Для трех групп обучаемых, участвовавших в исследовании, были включены различные возможности и ограничения в СДО. Для первой группы были доступны материалы всех типов, и обучаемые могли самостоятельно выбирать материал для изучения. Это контрольная группа, на которую не оказывалось никакого влияния. Для второй группы агент анализировал тип восприятия обучаемого и выдавал рекомендацию по выбору типа материала. Другие типы материалов были также доступны. Для третьей группы агент, как и для второй группы, анализировал тип восприятия студента, но при этом он сам решал, какой материал следует изучать. Доступ к другим типам материала ограничивался. Разбор задач повышенной сложности и усложненные задачи были доступны всем студентам без ограничений.

Работа агента заключалась в сборе предварительной информации о студентах (определение типа восприятия обучаемого, выявление психологического состояния и мотивации в начале дистанционного обучения) и в предложении ему лучшего варианта обучения на основании его психологического состояния (рис. 2).

В систему изначально был заложен функционал сбора статистики всех происходящих в ней событий. К таким событиям можно отнести входы пользователей в систему и выходы из нее, открытие тем, лекций и контрольных работ, прохождение тестов, изучение различных типов материала, время, потраченное на изучение материала. Помимо этого, ведется подробная статистика достигнутых студентами результатов: быстрота сдачи контрольной работы, количество пересдач, количество пройденных тестов, количество успешно пройденных тестов, время, затрачиваемое на тесты (в том числе среднее).

Для проведения психологического тестирования с требуемым количеством повторений требова-

лось максимально точно предугадать частоту использования студентами СДО и предложить им требуемое количество прохождений тестов, определяющих модальность.

Так как в распоряжении авторов была статистика входов в систему исследуемых студентов за несколько лет работы в СДО, простым вариантом казался следующий: время регистрации в СДО (в днях) разделить на количество входов в систему за этот период, результат покажет средний интервал между входами в систему.

К сожалению, у данного подхода выявились недостатки. Полученные значения были слишком низкими, так как у студентов заочной формы обучения велики перерывы в обучении, когда они не учатся, а также праздничные дни и каникулы. Требовалось исключить эти периоды из расчетов. Также не каждый вход в систему ведет к учебной деятельности, например, вход произведен, но никакое учебное действие не зафиксировано, поэтому целесообразно определить весовые коэффициенты, оценивающие вход в систему в том или ином случае (табл. 1).

Коэффициенты могут складываться в пределах одного входа, но суммарно не более 1.

Далее с учетом всех показателей была выведена формула для подсчета среднего интервала (в днях) между входами в СДО:

$$\tau = \frac{t}{\sum_{i=1}^n k_i}, \quad (1)$$

где t – суммарное количество учебных дней, проведенных в системе СДО с момента первого входа (в днях); n – количество входов; k – суммарный коэффициент.

По итогам сбора первичной статистики агент вычисляет, сколько раз он сможет провести тестирование.

В СДО уже присутствовала полноценная система проверки знаний на основе тестов, обслужи-

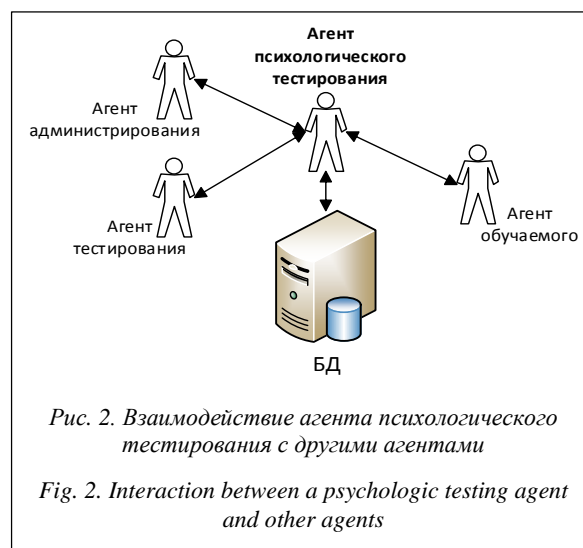


Таблица 1

Весовые коэффициенты входа в систему

Table 1

Log-in weight coefficients

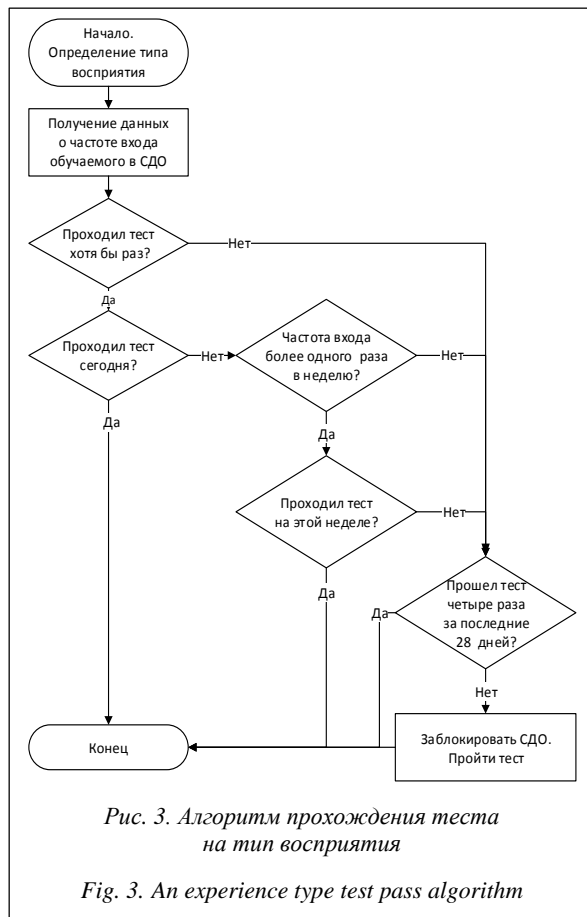
Показатель (за 1 сеанс)	Коэффициент	Обоснование
Вход в систему во внеучебное время	0,4	Вход в систему во внеучебное время нетипичен, поэтому стоит учитывать его отдельным показателем. Такие входы редки, так как в таком случае работают с СДО либо дополнительно, либо в целях пересдачи
Вход в систему	0,2	Сам по себе вход в систему мало что дает, но открывает путь к другим коэффициентам
Вход в систему без учебной деятельности	0	Так как учебная деятельность не была зафиксирована, такой вход считается «случайным». Например, открытие браузера, в котором в одной из вкладок закреплена СДО. В таком случае система автоматически загружается, но пользователь может даже не подозревать об этом
Посещение одной или нескольких лекций не менее 60 минут	0,4	Система проверяет активность пользователя каждые 3 минуты, поэтому может фиксировать время, которое он провел за изучением предметов. Если активность на странице прекратилась, СДО приостанавливает учет времени. Провести за изучением материала не менее одного часа – хорошая подготовительная работа
Посещение одной или нескольких лекций не менее 30 минут	0,3	За полчаса можно поверхностно ознакомиться с материалами лекции
Посещение одной или нескольких лекций не менее 10 минут	0,1	За 10 минут можно только повторить материал. Изучить и осмыслить за это время что-то новое проблематично
Принятие контрольной работы	0,2	Студент входит в новую для себя контрольную работу и получает ее задание. Для этого действия не требуется много времени, и само выполнение контрольной работы не всегда связано с СДО
Сдача/пересдача контрольной работы	0,2	Сдача контрольной работы подразумевает прикрепление и отправку документа в контрольной работе. Также не требует много времени
Полное решение двух и более тестов	0,4	Решение двух и более тестов требует достаточно много времени в СДО
Полное решение одного теста	0,3	Решение одного теста занимает до 30 минут
Начато решение теста	0,1	Тест не завершен. В СДО всегда можно вернуться и дорешать начатый тест (кроме тестов на время)
Закончено решение теста	0,1	Ранее начатый в СДО тест успешно завершен
Решение теста на положи- тельную оценку	0,1	Тест завершен с показателем, превышающим 50 %

ваемая агентом тестирования, поэтому было принято решение сделать психологическое тестирование одним из вариантов обычного тестирования. Интерфейс, элементы управления и структура БД были использованы повторно (http://www.swsys.ru/uploaded/image/2017_1/2017-1-dop/3.jpg).

Определение типа восприятия обучаемого проводилось методом предварительного прохождения психологического теста в системе. Студенты в течение 4 недель до начала эксперимента при входе в СДО получали тест на определение перцептивной модальности. Тест выдавался не чаще одного раза в день. Отказаться от его прохождения было

нельзя, дальнейшая работа в СДО блокировалась (http://www.swsys.ru/uploaded/image/2017_1/2017-1-dop/4.jpg).

Если агент считал, что вход осуществляется достаточно часто (не менее 1 раза в неделю), то предлагал тест один раз в неделю. Если же обучаемый заходил редко (менее 1 раза за 2 недели), агент использовал каждый возможный случай, чтобы заставить обучаемого пройти тест. Задача агента – провести не менее 2, но не более 5 тестов для каждого студента (рис. 3). В исследовании не ставилась задача подсчета погрешности агента в определении частоты входов.



После сбора статистической информации была включена основная функция агента – рекомендация вида материала на основании типа восприятия обучающегося.

Усредненная оценка по совокупности тестов принималась агентом в качестве входного параметра. Алгоритм расчета усредненного параметра включает несколько формул: простую, учитывающую только суммарные результаты проведенных тестов, и усложненную, когда простая формула не дает точного результата.

Предположим, в начале эксперимента у агента имелись данные, приведенные в таблице 2. В каждом отдельном тесте результат рассчитывается по формуле наибольшего соответствия психологическому типу:

$$f_i(x, y, z) = \begin{cases} V, & \text{если } (x > y) \text{ и } (x > z), \\ A, & \text{если } (y > x) \text{ и } (y > z), \\ K, & \text{если } (z > x) \text{ и } (z > y), \\ D, & \text{если } (x \approx y) \text{ и } (x \approx z) \\ & \text{и } (y \approx z) \text{ и } (\Delta_{x,y,z} \leq 2), \end{cases} \quad (2)$$

где V – результат модальности «визуал»; A – результат модальности «аудиал»; K – результат модальности «кинестетик»; D – результат модальности «дискрет»; x, y, z – сумма реакций, определяющих перцептивную модальность.

Рассмотрим случай первого студента – здесь во всех испытаниях преобладает оценка одной модальности, поэтому можно получить итоговую оценку по простой формуле

$$f_0 = \frac{\sum_{i=1}^n f(x, y, z)_i}{n}. \quad (3)$$

В случае со вторым студентом получены противоречивые данные, все контрольные измерения показали разные результаты, включая неопределенное состояние. Чтобы из этих данных получить результирующий показатель, обратимся к следующей формуле:

$$f_0 = \begin{cases} V, & \text{если } \left(\frac{\sum_{i=1}^n x_i}{n} > \frac{\sum_{i=1}^n y_i}{n} \right) \\ & \text{и } \left(\frac{\sum_{i=1}^n x_i}{n} > \frac{\sum_{i=1}^n z_i}{n} \right); \\ A, & \text{если } \left(\frac{\sum_{i=1}^n y_i}{n} > \frac{\sum_{i=1}^n x_i}{n} \right) \\ & \text{и } \left(\frac{\sum_{i=1}^n y_i}{n} > \frac{\sum_{i=1}^n z_i}{n} \right); \\ K, & \text{если } \left(\frac{\sum_{i=1}^n z_i}{n} > \frac{\sum_{i=1}^n x_i}{n} \right) \\ & \text{и } \left(\frac{\sum_{i=1}^n z_i}{n} > \frac{\sum_{i=1}^n y_i}{n} \right); \\ D, & \text{если } \left(\frac{\sum_{i=1}^n x_i}{n} \approx \frac{\sum_{i=1}^n y_i}{n} \right) \\ & \text{и } \left(\frac{\sum_{i=1}^n x_i}{n} \approx \frac{\sum_{i=1}^n z_i}{n} \right) \\ & \text{и } \left(\frac{\sum_{i=1}^n y_i}{n} \approx \frac{\sum_{i=1}^n z_i}{n} \right) \\ & \text{и } \left(\Delta \left(\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n y_i}{n}, \frac{\sum_{i=1}^n z_i}{n} \right) \leq 2 \right). \end{cases} \quad (4)$$

В результате агент при должном количестве измерений может давать более точный результат измерения перцептивной модальности (http://www.swsys.ru/uploaded/image/2017_1/2017-1-dop/8.jpg).

По итогам проведенного исследования сделан вывод о целесообразности внедрения агента психологического анализа в СДО (табл. 3).

Студенты, изучавшие подходящий им материал, лучше его усвоили и прошли контрольный тест. Дополнительным фактором, указывающим на

Таблица 2

Данные для определения итогового восприятия

Table 2

Data to define resulting perception

Испытуемый	Номер теста	Сумма реакций, определяющих перцептивную модальность			Результат
		Визуал	Аудиал	Кинестетик	
Студент 1	1	11	6	7	Визуал
	2	12	6	8	Визуал
	3	11	5	7	Визуал
	Итого	11,3	5,6	7,3	Визуал
Студент 2	1	4	10	8	Аудиал
	2	5	10	13	Кинестетик
	3	4	11	11	Дискрет
	Итого	4,3	10,3	10,6	Дискрет
Студент 3	1	7	7	8	Кинестетик
	2	9	8	8	Кинестетик
	3	13	8	8	Визуал
	Итого	9,6	7,6	8	Кинестетик

лучшее понимание основного материала, является самостоятельный переход студентов к изучению более сложных задач. Кроме того, эксперимент показал, что агент лучше помогает студентам, работая в рекомендательном режиме. Студенты, не имеющие возможности выбора, подготовились с более низким результатом, чем те, для кого система работала в рекомендательном режиме. С другой стороны, без использования рекомендаций студенты справились хуже, и в основном для изучения темы предпочтение отдавалось видеоматериалам (30 студентов из 33). Таким образом, не стоит применять запрет на выбор типа информации, так как это искусственно ограничивает обучаемых в их стремлении к знаниям.

Таблица 3

Результаты тестирования

Table 3

Test results

Группа	Прошли тестирование	Решение усложненной задачи
Контрольная группа (33 чел.)	42,4 % (14 чел.)	12,1 % (4 чел.)
Рекомендательный режим работы агента (39 чел.)	61,5 % (24 чел.)	51,3 % (20 чел.)
Обязывающий режим работы агента (36 чел.)	52,8 % (19 чел.)	41,7 % (15 чел.)

Исследование показало, что использование агента, определяющего тип восприятия студента, повышает эффективность СДО. Продолжение исследования направлено на повышение интеллектуальных свойств агента и поиск возможностей применения подобного подхода в системах корпоративного обучения.

Литература

1. Швецов А.Н., Сибирцев Е.В., Андрианов И.А. Компьютерные обучающие системы: мультиагентный подход // XII Всеросс. совещание по проблемам управления (ВСПУ-2014): сб. докл. Изд-во ИПУ РАН, 2014. С. 4048–4059.
2. Пашкин М.П. Многоагентная интеллектуальная система дистанционного обучения // Труды СПИИ РАН. 2006. Т. 1. Вып. 3. С. 126–137.
3. Тарасов В.Б. От многоагентных систем к интеллектуальным организациям: философия, психология, информатика. М.: Эдиториал УРСС, 2002. 352 с.
4. Андреев В.В., Виттих В.А., Батищев С.В. Методы и средства создания открытых мультиагентных систем для поддержки процессов принятия решений // Изв. РАН: Теория и системы управления. 2003. С. 126–137.
5. Бакаева А.С., Компаниец В.С. Сравнительный анализ психометрических тестов для автоматизированной диагностики перцептивной модальности обучающихся // Новое слово в науке: перспективы развития: матер. II Междунар. науч.-практ. конф. (30 дек. 2014 г., Чебоксары). Чебоксары: Интерактив плюс, 2014. С. 32–34.
6. Овчаренко О.М. Влияние перцептивной модальности на качество знаний и успеваемости. СПб: 2013. URL: <http://nsportal.ru/shkola/biologiya/library/2013/12/01/vliyanie-perseptivnoy-modalnosti-na-kachestvo-znaniy-i> (дата обращения: 10.07.2016).
7. Хрянин Е.Л. Эффективная организация образовательного процесса в вузе с использованием системы дистанционного обучения // Вузовская наука – региону: сб. матер. XII Всерос. науч.-технич. конф. Вологда: Изд-во Вологодского гос. ун-та, 2014. С. 69.
8. Стефанюк В.Л. Поведение многоагентных систем: парадигма координации // Новости искусственного интеллекта. 1997. № 4. С. 92–104.
9. Кузьмина Е.С., Свенцицкий А.Л. Промышленная социальная психология. Л.: Изд-во ЛГУ, 1982. С. 285.
10. Эйдемиллер Э., Юстицкис В. Психология и психотерапия семьи. СПб: Питер, 2008. С. 577–579.
11. Фетискин Н.П., Козлов В.В., Мануйлов Г.М. Социально-психологическая диагностика развития личности и малых групп. М.: Изд-во Ин-та психотерапии, 2002. 490 с.

A SOFTWARE AGENT TO DETERMINE STUDENT'S PSYCHOLOGICAL STATE IN E-LEARNING SYSTEMS

E.L. Khryanin^{1,2}, Postgraduate Student, Chief Engineer, evgeshah@list.ru

A.N. Shvetsov¹, Dr.Sc. (Engineering), Professor, smithv@mail.ru

¹ Vologda State University, Lenin St. 15, Vologda 160000, Russian Federation

² Bank "Vologzhanin", Batyushkov St. 11, Vologda, 160000, Russian Federation

Abstract. The article considers the problem of using software agents to assess students' psychological state in an e-learning system. The hypothesis of the study is the following: the more psychologically acceptable material for a student, the faster and better it is learned. It is required to develop an automatic algorithm for selection of material.

The article describes the developed e-learning system, which has been developed over 5 years and tested in one of the state universities. There is a brief description of e-learning system implementation that includes the agent interaction scheme, main database tables, backend and frontend implementation. The paper also describes a method and an algorithm to determine student's perceptual modality during psychological testing. It uses statistical methods to predict the probability of logging-in (based on statistics). The authors propose weight coefficients of frequency of using e-learning system by students for the agent, which determines their psychological state, to make decisions.

The paper describes the created algorithm of an automatic decision on the need in testing. The study involved 3 groups: a control group, a group with recommendation of material and a group with material chosen by an agent. The study involved more than 90 people. The study has formed formulas for perceptual modality calculation for several consecutive measurements. There is an example of calculation clarification for contradictory data. The experiment has shown positive results when using a recommendation mode. More than 61 % of students have passed the control test, and more than a half of the group has solved a difficult task (about 42 % and 12 % in the control group respectively).

There is a conclusion on expediency of using the psychological state definition agent in e-learning systems.

Keywords: agent-oriented approach, intelligent system, learning management system, psychological state diagnostics, perceptual modality, learning outcome assessment, PHP, MySQL, MVC.

References

1. Shvetsov A.N., Sibirtsev E.V., Andrianov I.A. Computer learning systems: multiagent approach. *XII Vseross. soveshchanie po problemam upravleniya VSPU-2014* [All-Russian Conf. on Management Problems VSPU-2014]. V.A. Trapeznikov Institute of Control Sciences of RAS Publ., 2014, pp. 4048–4059 (in Russ.).
2. Pashkin M.P. Multi-agent intelligent distance learning system. *Trudy SPIIRAN* [SPIIRAS Proc.]. RAN SPIIRAN Publ., St. Petersburg, 2006, vol. 1, no. 3, pp. 126–137.
3. Tarasov V.B. *Ot mnogoagentnykh sistem k intellektualnym organizatsiyam: filosofiya, psikhologiya, informatika* [From multiagent systems to intelligent companies: philosophy, psychology, computer science]. Moscow, Editorial URSS Publ., 2002.
4. Andreev V.V., Vittikh V.A., Batishchev S.V. Methods i means of creating open multiagent systems to support decision-making processes. *Izvestiya RAN. Teoriya i systemy upravleniya* [Journal of Computer and Systems Sciences Int.]. 2003, pp. 126–137 (in Russ.).
5. Bakaeva A.S., Kompaniets V.S. Comparative analysis of psychometric tests for automated diagnostics of students' perceptual modality. *Novoe slovo v nauke: perspektivy razvitiya: mater. II Mezhdunar. nauch.-praktich. konf.* [A New Word in Science: Development Prospects: Proc. 2nd Science and Practice Conf.]. Cheboksary, Interaktiv plyus Publ., 2014 (in Russ.).
6. Ovcharenko O.M. *Vliyanie pertseptivnoy modalnosti na kachestvo znaniy i uspevaemosti* [The impact of perceptual modality on the quality of knowledge and progress]. St. Petersburg, 2013. Available at: <http://nsportal.ru/shkola/biologiya/library/2013/12/01/vliyanie-pertseptivnoy-modalnosti-na-kachestvo-znaniy-i> (accessed July 10, 2016).
7. Khryanin E.L. Effective organization of an educational process in a university using a e-learning system. *Vuzovskaya nauka regionu: materialy XII Vseross. nauch.-tekhnich. konf.* [From Academic Science to a Region: Proc. 12th All-Russ. Science and Practice Conf.]. Vologda, Vologodsky gos. univ. Publ., 2014, p. 69 (in Russ.).
8. Stefanyuk V.L. The behavior of multiagent systems: coordination paradigm. *Novosti iskusstvennogo intellekta* [Artificial Intelligence News]. 1997, no. 4, pp. 92–104 (in Russ.).
9. Kuzmina E.S., Svetsitsky A.L. *Promyshlennaya sotsialnaya psikhologiya* [Industrial Social Psychology]. Leningrad, LGU Publ., 1982, 285 p.
10. Eydemiller E., Yustitskis V. *Psikhologiya i psikhoterapiya semi* [Family Psychology and Psychotherapy]. 4th ed., St. Petersburg, Piter Publ., 2008, 672 p.
11. Fetiskin N.P., Kozlov V.V., Manuylov G.M. *Sotsialno-psikhologicheskaya diagnostika razvitiya lichnosti i malyykh grupp* [Social-psychological diagnostics of personality and little group development]. Moscow, Institut Psikhoterapii Publ., 2002.

УДК 681.513.6

DOI: 10.15827/0236-235X.117.119-123

Дата подачи статьи: 08.12.16

2017. Т. 30. № 1. С. 119–123

МОНИТОРИНГ ЧАСТОТНОГО РЕСУРСА ГЕОСТАЦИОНАРНЫХ СПУТНИКОВ-РЕТРАНСЛЯТОРОВ С ИСПОЛЬЗОВАНИЕМ ЭНТРОПИИ ПОКРЫТИЯ

А.В. Сухов, д.т.н., профессор, avs57@mail.ru

(Московский авиационный институт (национальный исследовательский университет), Волоколамское шоссе, 4, г. Москва, 125993, Россия);

В.Н. Решетников, д.ф.-м.н., профессор, главный научный сотрудник, rvp_@mail.ru

(Центр визуализации и спутниковых информационных технологий НИИСИ РАН, Нахимовский просп., 36, корп. 1, г. Москва, 117218, Россия);

С.Б. Савилкин, к.ф.-м.н., доцент, Savilkin@mail.ru

(Московский авиационный институт (национальный исследовательский университет), Волоколамское шоссе, 4, г. Москва, 125993, Россия)

Рассмотрен мониторинг радиочастотного спектра для спутников-ретрансляторов, размещаемых на геостационарных орбитах. При этом решена оптимизационная задача обнаружения источников помех при заданных времени поиска и точности определения координат источников помех. Оптимизационная задача решена в целевом информационном пространстве, основанном на энтропии покрытия.

Определение местоположения наземных несанкционированных радиопередатчиков выполняется путем анализа временной задержки сигнала и доплеровского смещения частоты сигнала.

Местоположение источника помех на поверхности Земли можно определить, используя сигналы передатчика, ретранслируемые через одиночный спутник связи на геостационарной орбите. Небольшой доплеровский сдвиг несущей частоты сигнала, вызванный небольшим перемещением данного спутника на орбите относительно поверхности Земли, можно использовать для расчета местоположения передатчика.

В настоящей работе основное внимание сосредоточено на потенциальной точности оценок и выборе эффективного в смысле минимума энтропии покрытия подхода к оптимизации времени измерений.

Время сеанса проведения измерений, отношение сигнал/шум и параметры проведения измерений взаимосвязаны между собой. Отношения между реальными и нормативными параметрами измерений использованы в информационной мере – энтропии покрытия.

Энтропия покрытия характеризует эффективность систем, которые могут быть представлены вектором показателей эффективности, в соответствии с их целевым применением. Минимальное значение, равное нулю, означает, что нормативные требования выполнены полностью, а положительные значения характеризуют уровень обобщенного несоответствия нормативным требованиям.

С использованием энтропии покрытия проведена оценка информационной потенциальной эффективности при обнаружении координат источника помех с использованием эффекта доплеровского сдвига частоты.

Ключевые слова: мониторинг радиочастотного спектра, энтропия покрытия, геостационарная орбита, средства измерений, оценка.

Одной из актуальных задач мониторинга частотного ресурса, связанных с эффективным использованием спутников-ретрансляторов, расположенных на геостационарных орбитах (ГСО), является противодействие несанкционированному использованию, как правило, с хакерским взломом защиты. И тогда встает вопрос об идентификации передатчика. При этом расположенные на Земле источники помех используют спутниковый канал по линии вверх, а сигнал по линии вниз будет восприниматься как помеха.

Методы определения местоположения передатчиков на поверхности Земли для спутников на ГСО

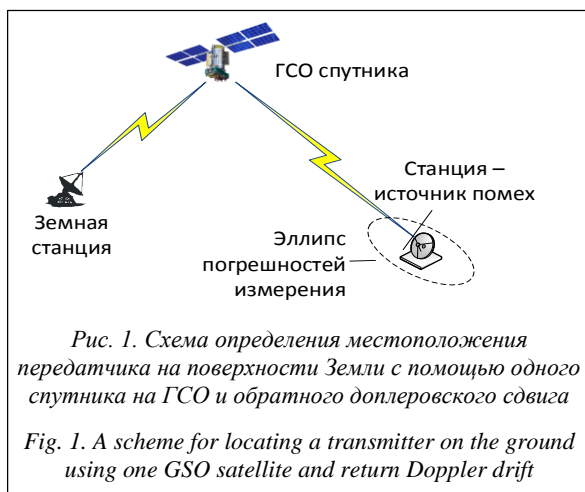
Определение местоположения несанкционированных наземных радиопередатчиков, затрагивающих спутники связи на ГСО, является сложной задачей, которая обычно выполняется путем анализа

временной задержки сигнала (ВЗС) и доплеровского смещения частоты (ДСЧ) сигнала. Общее описание этих подходов приведено в [1].

При проведении измерений для повышения точности определения координат источника помех может использоваться сигнал от второго спутника ГСО, если он попадает в луч передатчика станции помехи. Земных станций приема сигналов от спутников также может быть две и более, при этом земные станции должны быть синхронизированы.

Рассмотрим предельные показатели эффективности определения местоположения передатчиков на поверхности Земли с помощью одного спутника на ГСО и обратного доплеровского сдвига [1–6] (рис. 1).

Местоположение передатчика (или источника помех) на поверхности Земли можно определить при конкретных условиях, используя сигналы передатчика, ретранслируемые через одиночный спутник связи на ГСО. Небольшой доплеровский



сдвиг несущей частоты сигнала, вызванный небольшим перемещением данного спутника на орбите относительно поверхности Земли, можно использовать для расчета местоположения передатчика вплоть до указания определенной точки (в пределах десятков километров), где можно развернуть подвижные средства для точного установления местоположения передатчика или источника помех. В этой методике используются измерения несущей частоты нерегулярных кратковременных передач, разнесенных по времени на несколько часов. Для оценки с необходимой степенью точности малого доплеровского сдвига в передаваемых сигналах используются методы со сверхвысокой разрешающей способностью и высоким качеством обработки сигнала. Прогнозы позиций и скорости спутника могут уточняться при использовании эталонного передатчика [3].

Эксцентриситет и ненулевой наклон плоскости ГСО приводят к некоторому перемещению спутника относительно поверхности Земли. При таком перемещении создается небольшой доплеровский сдвиг, который можно использовать для расчета местоположения передатчика. В целях обеспечения необходимой точности определения координат источника помех могут использоваться наблюдения за излучениями передатчиков, разнесенными по времени на несколько часов. Чтобы можно было использовать этот метод, несущие частоты требуются измерять с очень высокой точностью.

В [1] приведено выражение для прогнозирования несущей частоты сигнала, который ретранслируется через спутник на ГСО:

$$f_R = \left[f_T \left(1 + \frac{v_s^T \cdot (r - r_s)}{c \cdot \|r - r_s\|} \right) + \Delta_f \right] \left(1 + \frac{v_D}{c} \right), \quad (1)$$

где f_R – несущая частота принимаемого сигнала; f_T – несущая частота передаваемого сигнала; v_s – вектор скорости спутника в период наблюдения; r_s – вектор позиции спутника в период наблюдения; r – вектор позиции передатчика; Δ_f – транспонирование частоты в ретрансляторе спутника; v_D – скалярная скорость изменения дальности между спут-

ником и приемником; c – скорость распространения сигнала.

Векторы r_s и r определены в пространственной геоцентрической декартовой системе координат.

Для определения местоположения передатчика на поверхности Земли в алгоритме [3] используется доплеровский сдвиг частоты, который возникает на линии вверх (от источника помех на спутник) и на линии вниз (от спутника на земную станцию).

Измерения частоты на приемнике земной станции осуществляются достаточно продолжительное время. Определение местоположения источника помех на поверхности Земли выполняется на основе получаемых оценок доплеровского смещения частоты, для чего следует использовать алгоритмы марковской теории оптимальной нелинейной фильтрации в дискретном времени [7]. Эти алгоритмы основаны на накоплении полезной информации о параметрах сигнала, которые содержатся в апостериорной плотности вероятности.

В настоящей работе основное внимание сосредоточено на потенциальной точности оценок и на выборе эффективного в смысле минимума энтропии покрытия [8] подхода.

В алгоритме для определения местоположения на поверхности Земли на основе выражения (1) применим линеаризованный вариант выражения принимаемой частоты, для чего используем разложение правой части выражения (1) в ряд Тейлора и ограничимся членами первого порядка малости. В результате приходим к системе линейных уравнений с представлением в виде матричного уравнения [1, 3–5]:

$$F = \Delta A, \quad (2)$$

где F – вектор совокупности k измерений несущей частоты принятого сигнала: $F^t = (f_1, f_2, \dots, f_k)$; t – знак транспонирования; $A^t = [A_f A_x A_y A_z]^t$, каждый вектор-столбец A является производной правой части выражения (1) по каждому параметру:

$$A_f = \frac{\partial \left\{ \left[f_T \left(1 + \frac{v_s^T \cdot (r - r_s)}{c \cdot \|r - r_s\|} \right) + \Delta_f \right] \left(1 + \frac{v_D}{c} \right) \right\}}{\partial f},$$

$$A_x = \frac{\partial \left\{ \left[f_T \left(1 + \frac{v_s^T \cdot (r - r_s)}{c \cdot \|r - r_s\|} \right) + \Delta_f \right] \left(1 + \frac{v_D}{c} \right) \right\}}{\partial x},$$

$$A_y = \frac{\partial \left\{ \left[f_T \left(1 + \frac{v_s^T \cdot (r - r_s)}{c \cdot \|r - r_s\|} \right) + \Delta_f \right] \left(1 + \frac{v_D}{c} \right) \right\}}{\partial y},$$

$$A_z = \frac{\partial \left\{ \left[f_T \left(1 + \frac{v_s^T \cdot (r - r_s)}{c \cdot \|r - r_s\|} \right) + \Delta_f \right] \left(1 + \frac{v_D}{c} \right) \right\}}{\partial z},$$

где Δ – матрица размером $k \times 4$ погрешностей между истинными и оценочными значениями параметров f, x, y, z на каждом из k моментов времени.

Данное матричное уравнение используется в качестве детерминированной составляющей в уравнении наблюдения для алгоритмов квазиоптимальной нелинейной фильтрации, использующих расширенный дискретный фильтр Калмана [8].

Точность результатов определения местоположения на поверхности Земли зависит от погрешности определения векторов позиции и скорости спутника во время каждого сеанса наблюдения. Векторы обычно вычисляются на основе модели движения по орбите и совокупности из шести орбитальных элементов, которые характеризуют орбиту спутника. Элементы периодически уточняются на основе наблюдений спутника и полученных совокупностей элементов орбиты. Алгоритмами движения по орбите моделируется влияние гравитационных полей Земли, Солнца и Луны для прогнозирования позиции и скорости спутника на более длительные периоды времени, чем то время, на которое эти элементы были вычислены [6].

Информационная оценка эффективности измерений с использованием энтропии покрытия

Выражение (1) в правой части содержит в качестве неизвестных параметров как вектор координат r , так и частоту передатчика f_T . Для оценки потенциальной точности целесообразно разнести эти параметры в разные части равенства:

$$\left[\frac{f_R}{1 + \frac{v_D}{c}} - \Delta_f \right] \frac{c}{f_T} - c = \frac{v_s^T (r - r_s)}{\|r - r_s\|}. \quad (3)$$

В выражении (3) значение c/f_T представим длиной радиоволны передатчика λ_T :

$$\left[\frac{f_R}{1 + \frac{v_D}{c}} - \Delta_f \right] \lambda_T - c = \frac{v_s^T (r - r_s)}{\|r - r_s\|}. \quad (4)$$

В левой части выражения (4) случайной величиной является параметр λ_T , а в правой части – вектор координат передатчика r . В алгоритме определения координат [3] применена линеаризация выражения (1) путем разложения его правой части в ряд Тейлора с оставлением членов первого порядка малости. Для правой части выражения (4) также следует применить разложение в ряд Тейлора по вектору r с оставлением членов первого порядка малости. Тогда будет справедливо допущение о подобности законов распределения для λ_T и r , поскольку λ_T будет линейной комбинацией от компонентов вектора r .

В качестве параметрического шума для частоты сигнала будем использовать аддитивный белый гауссовский шум. Тогда плотность распределения примет вид:

$$p(f) = \frac{1}{\sqrt{2\pi}\sigma_f} e^{-\frac{(f-m_f)^2}{2\sigma_f^2}}, \quad (5)$$

где σ_f – среднееквадратическое отклонение (СКО) значений частоты; m_f – математическое ожидание значения частоты с учетом доплеровского сдвига.

Поскольку параметр λ_T функционально связан с f_T , его плотность распределения как плотность распределения функции случайной величины может быть определена следующим образом:

$$p(f) = \frac{1}{\sqrt{2\pi}\sigma_f\lambda^2} e^{-\frac{\left(\frac{c}{\lambda} - m_f\right)^2}{2\sigma_f^2}}. \quad (6)$$

С использованием неравенства Крамера–Рао может быть получено значение СКО для потенциальных оценок измерения доплеровского смещения частоты:

$$\sigma_f = \frac{\sqrt{3}}{2\pi T_n \sqrt{q}}, \quad (7)$$

где $q = 2E/N_0$ – отношение сигнал/шум по энергии сигнала; T_n – интервал времени наблюдения сигнала.

Для уточненных элементов орбиты геостационарного спутника в примере, приведенном в [1], использованы значения частот от 248,934968 МГц до 248,934974 МГц. Доплеровское смещение частоты составляет примерно 6 Гц, $\sigma_f \approx 8,71728E-05$ Гц, $\sigma_\lambda \approx 1,83E+10$ м.

Будем использовать эти данные для проведения оценки эффективности измерений. В этом случае с учетом того, что погрешности определения дальности и измерения частоты пропорциональны, получаем, что при отношении сигнал/шум 10 дБ значение СКО по дальности составит $\sigma_D \approx 523,04$ м. График плотности распределения вероятностей для измерений длины волны передатчика представлен на рисунке 2.

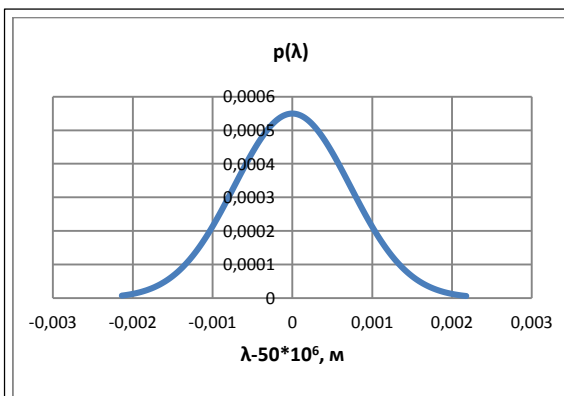


Рис. 2. Плотность распределения вероятностей для измерений длины волны передатчика

Fig. 2. Probability distribution frequency to measure a transmitter wave length

Из приведенного материала видно, что время сеанса проведения измерений, отношение сигнал/шум и параметры проведения измерений взаимосвязаны между собой. Возникает вопрос об оптимизации измерений. Ясно, что, чем дольше проводить измерения, тем меньше их погрешности. Но несанкционированные передачи могут быть как кратковременными (до нескольких секунд), так и достаточно продолжительными (до нескольких часов). Для информационной оценки эффективности необходимо задать тактико-технические требования по нормативной длительности наблюдения и нормативному СКО оценки дальности. Тогда можно использовать информационную меру – энтропию покрытия [7–9]: $H_{\Pi}(D_p) = k \log \{ \|C_{Dp} D_m \cup \cup D_m\|_2 / \|D_m\|_2 \}$, где C_{Dp} – операция дополнения к множеству реальных технических показателей D_p ; D_m – множество требуемых технических показателей; $\|\cdot\|_2$ – вторая норма. Коэффициент k определяет единицы измерения энтропии покрытия. Для двоичного логарифма энтропия покрытия измеряется в битах.

Энтропию покрытия в рассматриваемом случае доопределим следующим образом:

$$H_c = \log_2 \sqrt{\left(\frac{\begin{cases} T_n, T_n \geq T_{\text{норм}} \\ T_{\text{норм}}, T_n < T_{\text{норм}} \end{cases}}{T_{\text{норм}}} \right)^2 + \left(\frac{\begin{cases} \sigma_D, \sigma_D \geq \sigma_{\text{Днорм}} \\ \sigma_{\text{Днорм}}, \sigma_D < \sigma_{\text{Днорм}} \end{cases}}{\sigma_{\text{Днорм}}} \right)^2}, \quad (8)$$

где $T_{\text{норм}}$ и $\sigma_{\text{Днорм}}$ – нормативные значения времени наблюдения и СКО дальности соответственно.

Энтропия покрытия характеризует эффективность систем, которые могут быть представлены вектором показателей эффективности, в соответствии с их целевым применением. Минимальное значение, равное нулю, означает, что нормативные требования выполнены полностью, а положительные значения характеризуют уровень обобщенного несоответствия нормативным требованиям.

С использованием энтропии покрытия была проведена оценка информационной потенциальной эффективности при обнаружении координат источника помех с использованием эффекта доплеровского сдвига частоты. В качестве нормативного значения времени наблюдения использовано значение 10 сек., а в качестве нормативно допустимой погрешности измерения дальности использованы значения 100 и 1 000 метров. Результаты компьютерного моделирования при различных значениях сигнал/шум представлены на рисунках 3 и 4.

По результатам компьютерного моделирования можно сделать вывод, что оптимальный минимум энтропии покрытия связан со значением сигнал/шум в канале приема. При этом для нормативного СКО определения дальности 100 метров оптимальное время наблюдения должно составлять от 100 до 400 секунд, а для СКО определения дальности 1 000 метров – от 55 до 100 секунд.

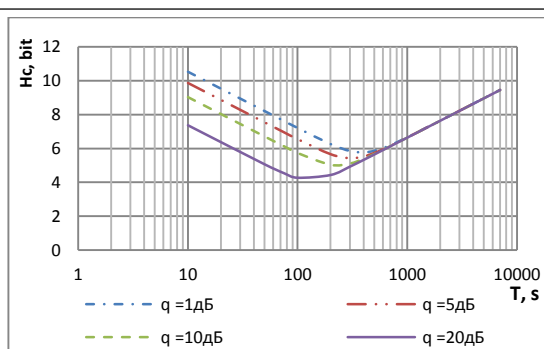


Рис. 3. Информационная эффективность оценки параметров источника помех для нормативного значения СКО дальности 100 метров

Fig. 3. Information efficiency of jamming source parameter estimation for normal value RMSD with the distance 100 meters

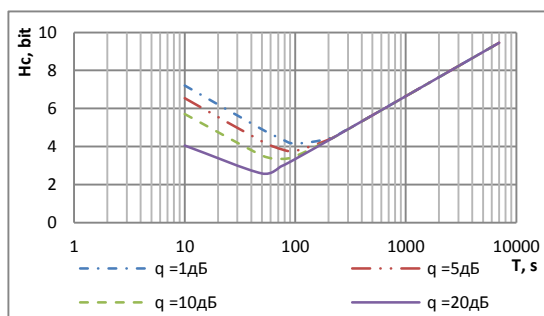


Рис. 4. Информационная эффективность оценки параметров источника помех для нормативного значения СКО дальности 1 000 метров

Fig. 4. Information efficiency of jamming source parameter estimation for normal value RMSD with the distance 1 000 meters

Энтропия покрытия позволяет не допускать лишних затрат на поиски источника помех, когда заданы требуемые точностные характеристики.

Работа выполнена при поддержке РФФИ, грант № 16-07-00511.

Литература

1. Handbook on Spectrum Monitoring. ITU, Switzerland, Geneva, 2011, pp. 374–387.
2. Черняк В.С. Многопозиционная радиолокация. М.: Радио и связь, 1993. 416 с.
3. Koets M.A., Bentley R.T. Satellite based geolocation using a single geosynchronous satellite and an inverse doppler technique. SRI Publ., USA, S.A., 1999.
4. Ковалев А.Н., Ковалев Ф.Н. Потенциальная точность определения параметров траектории цели в просветном бистатическом радиолокаторе // Тр. НГТУ им. П.Е. Алексеева. 2014. Т. 104. № 2. С. 17–22.
5. Рындык А.Г., Сидоров С.Б., Бляхман А.Б., Ковалев Ф.Н. Точность определения координат методом максимального правдоподобия при локации «на просвет» // Радиотехника и электроника. 1999. Т. 44. № 12. С. 1436–1440.
6. Алипбаев К.А. Моделирование движения геостационарного спутника // Вестн. КазНТУ. 2011. № 6. С. 120–122.
7. Ломакин М.И., Коровайцев А.А., Сухов А.В. Информа-

ционно-энтропийный подход к оценке метрологического ресурса средств измерений // Измерительная техника. 2014. № 6. С. 14–17.

8. Сухов А.В. Динамика информационных потоков в системе управления сложным техническим комплексом // Теория и системы управления. 2000. № 4. С. 111–120.

Software & Systems
DOI: 10.15827/0236-235X.117.119-123

Received 08.12.16
2017, vol. 30, no. 1, pp. 119–123

MONITORING OF FREQUENCY RESOURCE OF GEOSTATIONARY REPEATER SATELLITES USING COVER ENTROPY

A.V. Sukhov¹, Dr.Sc. (Engineering), Professor, avs57@mail.ru

V.N. Reshetnikov², Dr.Sc. (Physics and Mathematics), Professor, Chief Researcher, rvn_@mail.ru

S.B. Savilkin¹, Ph.D. (Physics and Mathematics), Associate Professor, Savilkin@mail.ru

¹ Moscow Aviation Institute (National Research University),
Volokolamskoe Highway 4, Moscow, 125993, Russian Federation

² Center of Visualization and Satellite Information Technologies SRISA,
Nakhimovsky Ave. 36/1, Moscow, 117218, Russian Federation

Abstract. The paper considers radio-frequency spectrum monitoring for repeater satellites placed in geostationary orbits. It solves the optimization problem of interference source detection with the given search time and definition accuracy of interference source coordinates. The optimization problem is solved in the target information space based on covering entropy.

Positioning of ground unauthorized radio transmitters is performed by analyzing a signal time delay and Doppler shift signal frequency.

The location of an interference source on the Earth's surface can be determined by transmitter signals, which are relayed through a single communications satellite to a geostationary orbit. A small Doppler shift of the signal carrier frequency, which is caused by a small displacement of a satellite on the orbit against the Earth's surface, can be used to calculate the transmitter location.

The paper focuses on potential accuracy of estimation and the choice of an efficient approach (in the sense of the minimum covering entropy) to optimization of measurement time.

Measurement session time, a signal-to-noise ratio and measurement parameters are interrelated. The relationships between real and specified parameters of measurements are used in information measure which is covering entropy (A. Sukhova).

The covering entropy characterizes the efficiency of systems that can be represented by a vector of performance indicators in accordance with their intended use. The minimum value of zero means that regulatory requirements are fulfilled, the positive values characterize the level of generalized compliance.

The authors evaluated information potential efficiency in detection of interference source coordinates using the Doppler frequency shift effect based on covering entropy.

Keywords: spectrum monitoring, covering entropy, geostationary orbit, measuring tools, evaluation.

Acknowledgements. The research has been financially supported by RFBR, grant no. 16-07-00511.

References

1. *Handbook on Spectrum Monitoring*. 2011 ed. ITU Publ., Switzerland, Geneva, 2011.
2. Chernyak V.S. *Mnogopozitsionnaya radiolokatsiya* [Multiposition Radiolocation]. Moscow, Radio i svyaz Publ., 1993, 416 p.
3. Koets M.A., Bentley R.T. *Satellite Based Geolocation Using a Single Geosynchronous Satellite and an Inverse Doppler Technique*. SRI Publ., USA, S.A., 1999.
4. Kovalev A.N., Kovalev F.N. Potential determination accuracy of target trajectory parameters in luminal prosvetnom bistatic radar. *Trudy Nizhegorodskogo gos. tekhnich. univ. im. R.E. Alekseeva* [Transactions of NNSTU n.a. R.E. Alekseev]. 2014, vol. 104, no. 2, pp. 17–22 (in Russ.).
5. Ryndyk A.G., Sidorov S.B., Blyakhman A.B., Kovalev F.N. Accuracy of position determination by the maximum likelihood method when locating peek-a-boo. *Radiotekhnika i elektronika* [Jour. of Communications Technology and Electronics]. 1999, vol. 44, no. 12, pp. 1436–1440 (in Russ.).
6. Alipbaev K.A. Geostationary satellite motion simulation. *Vestnik KazNTU* [Vestnik of KazNRTU]. 2011, no. 6, pp. 120–122 (in Russ.).
7. Lomakin M.I., Korovaytsev A.A., Sukhov A.V. Information-entropic approach to estimation of measuring device metrological resource. *Izmeritel'naya tekhnika* [Measurement Techniques]. 2014, no. 6, pp. 14–17 (in Russ.).
8. Sukhov A.V. Data flow dynamics in a complex technical management system. *Teoriya i sistemy upravleniya* [Jour. of Computer and Systems Sciences International]. 2000, no. 4, pp. 111–120 (in Russ.).

УДК 620.1

DOI: 10.15827/0236-235X.117.124-129

Дата подачи статьи: 11.07.16

2017. Т. 30. № 1. С. 124–129

СТАТИСТИЧЕСКИЙ АНАЛИЗ РЕЗУЛЬТАТОВ ИСПЫТАНИЙ ИЗДЕЛИЙ АВИАЦИОННОЙ ТЕХНИКИ В УСЛОВИЯХ СЛУЧАЙНОГО ЦЕНЗУРИРОВАНИЯ

А.В. Агамиров, д.т.н., профессор, tmk@mati.ru

(Национальный исследовательский университет «Московский энергетический институт»,
ул. Красноказарменная, 14, г. Москва, 111250, Россия);

В.А. Агамиров, к.т.н., avl095@mail.ru;

Вестяк В.А., к.ф.-м.н., доцент, kaf311@yandex.ru

(Московский авиационный институт (национальный исследовательский университет),
Волоколамское шоссе, 4, г. Москва, 125993, Россия)

В работе рассматривается методика точечного и интервального оценивания параметров распределений, применяемых при статистическом анализе усталостных испытаний элементов авиационных конструкций на базе метода наименьших квадратов, учитывающая наличие цензурированных наблюдений.

Актуальность работы определяется тем, что при решении задачи оценивания параметров распределений характеристик усталостных свойств для статистического анализа результатов усталостных испытаний изделий авиационной техники необходимо учитывать образцы, для которых произошла остановка испытаний до достижения ими критического состояния. Решение данной задачи с использованием известных методов (метода максимального правдоподобия) затруднено из-за немонотонности целевых функций, наличия ряда локальных экстремумов и т.д.

Первая часть статьи посвящена методике оценивания параметров распределения наблюдаемой случайной величины для полной выборки, преобразованной из многократно цензурированной (неполной) выборки путем бутстреп-моделирования, основанного на порядковых статистиках. Преобразование исходной случайно цензурированной выборки в квазиполную осуществляется для того, чтобы можно было использовать метод наименьших квадратов для оценки параметров распределения, поскольку этот метод применим только для полных выборок.

Во второй части статьи говорится о построении доверительных границ для квантиля распределения наблюдаемой случайной величины. В авиационной технике это применимо для оценки гарантированного ресурса, нормируемого по нижней доверительной границе квантиля долговечности.

Авторами разработана методика приведения в общем случае многократно цензурированной неполной выборки к эквивалентной квазиполной выборке, для которой можно использовать метод наименьших квадратов, а следовательно, получить наиболее устойчивые и эффективные оценки с минимальной дисперсией. Таким образом, решена задача точечного и интервального оценивания параметров распределений характеристик усталостных свойств элементов авиационных конструкций с учетом наличия многократно цензурированных наблюдений.

Ключевые слова: *живучесть, бутстреп-моделирование, метод наименьших квадратов, порядковые статистики, случайное цензурирование.*

Сокращение сроков ввода в эксплуатацию изделий авиационной техники, малые объемы подвергаемых испытаниям элементов авиационных конструкций, образующиеся в результате этих факторов незавершенные выборочные совокупности наблюдений требуют разработки и внедрения соответствующих методов статистического анализа, учитывающих эти обстоятельства. Многократно цензурированные выборки [1, 2] образуются в результате определения наработки ответственных элементов конструкции, достигших или не достигших критического состояния к моменту технического осмотра. Случайно цензурированные выборки могут образовываться также при испытаниях на усталость материалов и элементов двигателей, агрегатов и планера летательных аппаратов, когда ряд объектов в связи с ограничением времени не достигают критического состояния и снимаются с испытаний. В этих условиях одной из проблем является оценка характеристик долговечности и ресурса авиационных конструкций. Известные методы оценивания, такие как метод максимального правдоподобия [1–4], приводят к сложным систе-

мам нелинейных уравнений, решение которых затруднено из-за немонотонности целевых функций, наличия ряда локальных экстремумов и т.д.

В связи с этим в настоящей работе рассматривается задача моделирования и оценивания параметров незавершенных выборочных совокупностей методом наименьших квадратов (МНК), обладающим уникальными свойствами наиболее устойчивых и эффективных для малых выборок оценок с минимальной дисперсией.

В соответствии с МНК [3, 4] вектор оценок параметров линейной модели

$$y = X \cdot b, \quad (1)$$

где y – вектор-столбец наблюдений размерности n ; X – матрица размерности $n \times k_1$ известных коэффициентов ($n > k_1$); b – вектор-столбец параметров размерности k_1 , определяется из уравнения

$$\hat{b} = (X^T \cdot V^{-1} \cdot X)^{-1} \cdot X^T \cdot V^{-1} \cdot y. \quad (2)$$

Матрица рассеяния оценок b определяется из уравнения

$$D(\hat{b}) = (v) = \frac{\sigma^2}{n} (v^*) ; (v^*) = n \cdot (X^T \cdot V^{-1} \cdot X)^{-1}, \quad (3)$$

несмещенная оценка для остаточной дисперсии σ^2 определяется формулой

$$\hat{\sigma}^2 = s^2 = \frac{1}{n - k_1} \cdot (y - X \cdot \hat{b})^T \cdot V^{-1} \cdot (y - X \cdot \hat{b}), \quad (4)$$

где V – ковариационная матрица размерности $n \times n$ оценок параметров линейной модели.

Адекватность модели проверяется обычным способом на основании F -распределения дисперсионного отношения [3, 4].

Уравнения (3) и (4) позволяют оценивать параметры расположения (сдвига) и масштаба на основании порядковых статистик, то есть выборочных наблюдений, упорядоченных по величине. Пусть y_i – порядковые статистики, a и σ – параметры сдвига и масштаба (необязательно среднее и стандартное отклонения). В соответствии с МНК матрица X размерности $n \times 2$ имеет следующий вид:

$$X = \begin{pmatrix} 1 & \alpha_1 \\ 1 & \alpha_2 \\ \dots & \dots \\ 1_n & \alpha_n \end{pmatrix}, \quad (5)$$

где α – вектор-столбец размерности n математических ожиданий порядковых статистик.

Оценки параметров сдвига и масштаба и их матрица рассеяния определяются по уравнениям (2) и (3), где V – ковариационная матрица размерности $n \times n$ нормированных порядковых статистик.

Элементы вектора математических ожиданий (α) и ковариационной матрицы (V) нормированных порядковых статистик [3, 4] определяются из следующих уравнений:

$$\alpha_l = \frac{\int_{-\infty}^{\infty} x \cdot f(x) \cdot [1 - F(x)]^{n-l} \cdot [F(x)]^{l-1} \cdot dx}{B(l, n-l+1)}, \quad (6)$$

$$V_{l,l} = \frac{\int_{-\infty}^{\infty} x^2 \cdot f(x) \cdot [1 - F(x)]^{n-l} \cdot [F(x)]^{l-1} \cdot dx}{B(l, n-l+1)} - \alpha_l^2, \quad (7)$$

$$V_{l,s} (l < s) = \frac{\int_{-\infty}^{\infty} x \cdot f(x) \cdot [1 - F(x)]^{n-s} \cdot dx}{B(l, s-l+1) \cdot B(s, n-s+1)} \times$$

$$\times \frac{\int_{-\infty}^{\infty} y \cdot f(y) \cdot [F(x) - F(y)]^{s-l-1} \cdot [F(y)]^{l-1} \cdot dy}{B(l, s-l+1) \cdot B(s, n-s+1)} - \alpha_l \cdot \alpha_s, \quad (8)$$

$$B(a, b) = \frac{(b-1)! \cdot (a-1)!}{(a+b-1)!}, \quad (9)$$

где $l, s = 1, \dots, n$; $f(z)$, $F(z)$ – плотность и функция нормированного непрерывного распределения с параметрами сдвига и масштаба.

Для двухпараметрического логарифмически нормального $y = \ln x$ и нормального $y = x$ распределений

$$f(z) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{z^2}{2}}, \quad F(z) = \int_{-\infty}^z f(t) \cdot dt, \quad z = (y-a)/\sigma.$$

Для представления трехпараметрического распределения Вейбулла–Гнеденко $F(x) = 1 - e^{-\left(\frac{x-x_0}{c}\right)^b}$ к виду с параметрами сдвига и масштаба осуществляют следующее нормирующее преобразование:

$$y = \ln(x - x_0) = a + z \cdot \sigma, \quad \sigma = 1/b, \quad a = \ln c, \quad (10)$$

$$z = \ln \ln \frac{1}{1-F(z)}, \quad F(z) = 1 - e^{-e^z}, \quad f(z) = e^{z-e^z}.$$

В случае трехпараметрических логарифмически нормального и распределения Вейбулла–Гнеденко $y = \ln(x - x_0)$, где x_0 – независимая оценка порогового значения случайной величины или $x_0 = 0$.

Для однократно цензурированной справа выборки оценки параметров сдвига и масштаба и их ковариационная матрица определяются по тем же формулам, но при этом матрица X , вектор наблюдений y , ковариационная матрица V составляются по первым k наблюдениям случайной величины из n объектов, подвергшихся испытанию, а величина n в вышеприведенных формулах остается неизменной.

Интервальные оценки квантиля распределения для полной выборки из нормального или логарифмически нормального закона распределения (см. [5, 6]) определяются уравнениями:

$$\hat{x}_{pl} = \hat{a} + t_{1-\beta} \left[n-1, z_p \cdot \sqrt{n} \right] \cdot \frac{\hat{\sigma}}{\sqrt{n}}, \quad (11)$$

$$\hat{x}_{pu} = \hat{a} + t_{\beta} \left[n-1, z_p \cdot \sqrt{n} \right] \cdot \frac{\hat{\sigma}}{\sqrt{n}}, \quad (12)$$

где x_{pl} , x_{pu} – нижняя и верхняя доверительные границы для квантиля распределения x_p уровня вероятности P ; β – уровень доверительной вероятности (обычно $\beta=0,9$ или $0,95$); $t_{\gamma}[f, \Delta]$ – квантиль уровня γ нецентрального распределения Стьюдента с $f = n-1$ степенями свободы и с параметром нецентральности $\Delta = z_p \cdot \sqrt{n}$; z_p – квантиль уровня P нормированного нормального распределения; \hat{a} , $\hat{\sigma}$ – оценки параметров нормального распределения.

Точное значение квантиля нецентрального распределения Стьюдента t определяется по таблицам или в соответствии с разработанными вычислительными алгоритмами [6].

Для распределения Вейбулла–Гнеденко с параметрами сдвига и масштаба, а также в цензурированных выборках могут быть вычислены приближенные доверительные интервалы для квантилей распределения на основании нормальной аппроксимации. С этой целью предположим приближенно нормальным закон распределения случайной величины

$$\varphi = x_p - \hat{x}_{pu} = x_p - \hat{a} - t \cdot \frac{\hat{\sigma}}{\sqrt{n}}$$

с математическим ожиданием

$$M\{\varphi\} \approx x_p - a - t \cdot \frac{\sigma}{\sqrt{n}} \quad (13)$$

и дисперсией

$$D\{\varphi\} \approx D\{\hat{a}\} + 2 \cdot t \cdot \frac{\sigma \cdot D\{\hat{a}, \hat{\sigma}\}}{\sqrt{n}} + t^2 \cdot \frac{D\{\hat{\sigma}\}}{n}, \quad (14)$$

где $x_p = a + z_p \cdot \sigma$ – квантиль распределения. Элементы ковариационной матрицы (v) оценок параметров в соответствии с (3)

$$(v) = \frac{\sigma^2}{n} \cdot (v^*) = \begin{pmatrix} D\{\hat{a}\} & D\{\hat{a}, \hat{\sigma}\} \\ D\{\hat{\sigma}, \hat{a}\} & D\{\hat{\sigma}\} \end{pmatrix}. \quad (15)$$

$$D\{\hat{a}\} = \frac{\sigma^2}{n} \cdot v_{1,1}^*, \quad D\{\hat{\sigma}\} = \frac{\sigma^2}{n} \cdot v_{2,2}^*,$$

$$D\{\hat{a}, \hat{\sigma}\} = \frac{\sigma^2}{n} \cdot v_{1,2}^*. \quad (16)$$

Вероятность β того, что $P\{\varphi < 0\}$, приводит к следующему приближенному уравнению:

$$z_\beta \approx \frac{M\{\varphi\}}{\sqrt{D\{\varphi\}}} = \frac{z_p \cdot \sqrt{n} - t}{\sqrt{v_{1,1}^* + 2 \cdot \frac{t}{\sqrt{n}} \cdot v_{1,2}^* + \frac{t^2}{n} \cdot v_{2,2}^*}}, \quad (17)$$

где z_β – квантиль уровня β нормированного нормального распределения.

После преобразований определяется приближенное значение t , соответствующее числу степеней свободы $f = n - 1$, параметру нецентральности $\Delta = z_p \cdot \sqrt{n}$ и доверительной вероятности β :

$$t = t_{\beta, 1-\beta}[\Delta, f = n - 1] = \frac{A \pm \sqrt{A^2 - B \cdot C}}{B}, \quad (18)$$

$$A = \left[1 - \frac{v_{2,2}^*}{2 \cdot f} \right] \cdot \Delta - \frac{z_\beta^2 \cdot v_{1,2}^*}{\sqrt{n}},$$

$$B = \left[1 - \frac{v_{2,2}^*}{2 \cdot f} \right]^2 - \frac{z_\beta^2 \cdot v_{2,2}^*}{f}, \quad (19)$$

$$C = \Delta^2 - z_\beta^2 \cdot v_{1,1}^*.$$

В формулах (19) учтены поправки на смещение оценок, имеющие место в полной выборке. По формуле (18) осуществляется аппроксимация нецентрального распределения Стьюдента для полной выборки, при этом $v_{1,1}^* = 1$, $v_{2,2}^* = 0,5$, $v_{1,2}^* = 0$:

$$t_{\beta, 1-\beta}[\Delta, f = n - 1] \approx \frac{\left(1 - \frac{1}{4 \cdot f} \right) \cdot \Delta \pm \sqrt{\left(1 - \frac{1}{4 \cdot f} \right)^2 - \frac{z_\beta^2}{2 \cdot f} + \frac{\Delta^2}{2 \cdot f}}}{\left(1 - \frac{1}{4 \cdot f} \right)^2 - \frac{z_\beta^2}{2 \cdot f}}. \quad (20)$$

Доверительные границы для параметра сдвига a получают из (11) и (12) как частный случай при $z_p = 0$, $\Delta = 0$. Для нормального закона эти границы совпадают с доверительными границами для медианы распределения. В этом случае нецентральное распределение Стьюдента вырождается в хорошо табулированное центральное t -распределение Стьюдента. Для нормального распределения пара-

метр нецентральности определяется по формуле $\Delta = z_p \cdot \sqrt{n}$ (z_p – квантиль нормированного нормального распределения). Для распределения Вейбулла–Гнеденко, представленного в виде распределения с параметрами сдвига и масштаба, как показано выше, параметр нецентральности определяется из уравнения $\Delta = z_p \cdot \sqrt{n} = \ln \ln \frac{1}{1-p} \cdot \sqrt{n}$,

где $p = F(z_p) = 1 - e^{-e^{z_p}}$.

Необходимо отметить, что точные доверительные границы (11), (12) для квантиля распределения случайной величины и их аппроксимации (19), (20) получены в предположении вариации параметров сдвига и масштаба. Однако, как отмечается в работах [7, 8], тогда эти границы могут оказаться неоправданно широкими при малых объемах наблюдений, свойственных испытаниям авиационных конструкций, что приводит к весьма заниженным оценкам гарантированного ресурса, нормируемого по нижней доверительной границе квантиля долговечности. В таком случае часто пренебрегают вариацией параметра масштаба $\hat{\sigma}$, заменяя его априорным значением σ , полученным по результатам большого числа предварительных испытаний (для конструкций планера, например, среднее квадратичное отклонение логарифма долговечности $\sigma_{\lg N}$ предполагается равным 0,15). В этом случае матрица X (5) представляет собой вектор из единиц размерности n , а в уравнениях (19) следует положить равными нулю все элементы матрицы рассеяния, имеющие индексы 2,2 и 1,2:

$t_{\beta, 1-\beta}[\Delta, f = n - 1] \approx \Delta + z_{\beta, 1-\beta} \cdot \sqrt{v_{1,1}^*}$. Для полной выборки уравнение (20) примет следующий вид:

$$t_{\beta, 1-\beta}[\Delta, f = n - 1] \approx \Delta + z_{\beta, 1-\beta}.$$

Для применения рассмотренных выше математических моделей в случайно цензурированных выборках в настоящей работе предлагается формирование эквивалентной (квазиполной) выборки на базе исходной цензурированной путем бутстреп-моделирования случайных чисел в диапазоне, ограниченном наблюдаемыми порядковыми статистиками. С этой целью методом Монте-Карло моделировалась полная выборка (на основе нормального закона распределения и распределения Вейбулла–Гнеденко), в которой случайным образом путем моделирования равномерно распределенных случайных чисел в заданном диапазоне формировались объекты, не достигшие критического состояния. Доля таких объектов составляла от 0 до 50 % при объемах выборки 10, 12, 15 и 20. Цензурированные наблюдения заменялись далее случайно выбранными результатами наблюдений той же выборки, достигшими критического состояния. В дальнейшем выборка сортируется и формируется эквивалентная квазиполная выборка, параметры

которой могут быть определены как обычным непараметрическим методом:

$$\begin{aligned}\hat{a} &= \sum_{i=1}^n y_i / n, \hat{\sigma}^2 = \\ &= \sum_{i=1}^n (y_i - \bar{a})^2 / (n-1), D\{\hat{a}\} \approx \\ &\approx \hat{\sigma}^2 / n, D\{\hat{\sigma}\} \approx \hat{\sigma}^2 / [2 \cdot (n-1)],\end{aligned}\quad (21)$$

так и рассмотренным выше МНК, имеющим лучшие показатели эффективности.

Для проверки предлагаемой модели проводилось статистическое моделирование с многократным (до 2 000 раз) повторением испытаний, в каждом из которых оценивались параметры эквивалентных выборок по уравнению (2) и доверительная вероятность накрытия границами (11), (12) квантилей распределения. Для определения математических ожиданий и ковариаций порядковых статистик использовалась авторская вычислительная программа, основанная на разложении в ряд Корниша–Фишера. Расчетная двусторонняя доверительная вероятность β составляла 0,9. Уровень квантиля распределения P задавался 0,01. Эти результаты сравнивались с оценками, полученными для полной выборки. Результаты расчетов показали хорошее соответствие результатов для параметров сдвига и масштаба в пределах 1–2 % относительных погрешностей. Несколько большая погрешность (до 5 % с ростом степени цензурирования) наблюдалась в оценке доверительных вероятностей для квантилей распределения. Некоторые результаты моделирования, иллюстрирующие эти данные, представлены в таблице 1 для объемов испытаний 10 и 20. Расчеты проводились также для распределения Вейбулла–Гнеденко с близкими по точности результатами. Сравнение с непараметрическими оценками (21) показало незначительное отличие в точечных оценках параметров, но существенно меньшую дисперсию оценок для МНК, как и следовало ожидать.

Аналогичный вычислительный алгоритм может быть применен для задачи оценивания параметров реальной случайно цензурированной выборки, его блок-схема показана на рисунке 1. Отметим, что при моделировании случайных чисел в диапазоне порядковых статистик предполагается, что моделируемое наблюдение не может быть меньше соответствующего цензурированного значения. С этой целью в блок-схеме предусмотрен цикл возврата при невыполнении данного условия. Программы расчета параметров случайно цензурированных выборок на языке Javascript существуют в открытом доступе на сайте <http://inteh.mpei.ru>.

Пример 1. В таблице 2 представлены результаты усталостных испытаний на изгиб 20 лопаток из титанового сплава компрессора низкого давления авиационного двигателя при симметричном цикле амплитуды переменных напряжений. Звез-

дочками обозначены значения логарифмов долговечностей лопаток, не достигших критического состояния к моменту снятия с испытаний. Произвести оценку параметров нормального распределения логарифма долговечности и вычислить приближенные 90 %-ные доверительные границы для квантиля уровня $P=0,01$.

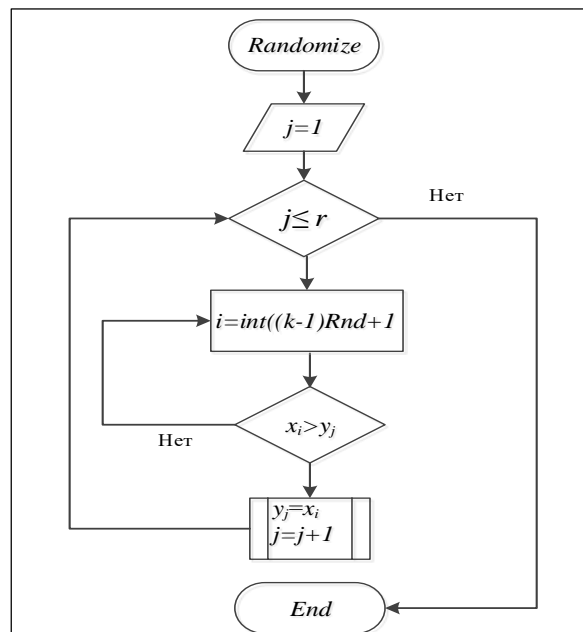


Рис. 1. Блок-схема алгоритма моделирования случайно цензурированной выборки:

n – объем выборки; k – объем наблюдений; r – количество цензурированных элементов выборки ($n=k+r$); Rnd – равномерно распределенное случайное число в диапазоне от 0 до 1; x_i – вектор значений наблюдений размерности k ($i=1, \dots, k$); y_j – вектор значений цензурированных наблюдений размерности r ($j=1, \dots, r$)

Fig. 1. A control flow chart of modeling a randomly censored sample

В таблице 3 представлена та же выборка, но обработанная с помощью бутстреп-моделирования цензурированных элементов в соответствии с алгоритмом, представленным на рисунке 1.

Оценки параметров нормального закона распределения логарифма долговечности \hat{a}_{lgN} , $\hat{\sigma}_{lgN}$ вычислялись методом наименьших квадратов по уравнению (2), оценки дисперсий параметров – по уравнению (3), доверительные границы \hat{x}_{pl} , \hat{x}_{pu} – по формулам (11), (12) с учетом (18), (19). Оценка квантиля уровня 0,01 определялась по формуле $\hat{x}_p = \hat{a}_{lgN} - 2,326 \cdot \hat{\sigma}_{lgN}$. Результаты расчетов представлены в таблице 4.

Функция распределения логарифма долговечности представлена на рисунке 2 на нормальной вероятностной бумаге. Там же отмечены опытные значения по таблице 3.

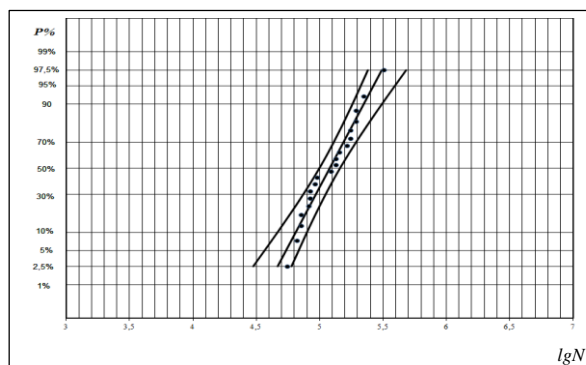


Рис. 2. Эмпирическая функция распределения логарифма долговечности на нормальной вероятностной бумаге

Fig. 2. Lifetime logarithm empirical distribution function on a normal probability paper

На основании изложенного сделаем следующие выводы.

В работе рассмотрена методика точечного и интервального оценивания параметров распределений, применяемых при статистическом анализе усталостных испытаний элементов авиационных конструкций на базе метода наименьших квадратов, учитывающая наличие цензурированных наблюдений.

С целью адаптации методики для случайно цензурированных выборок разработаны модель и алгоритм моделирования порядковых статистик, которые позволяют получить эквивалентную (квазиполную) выборку.

Для проверки методики проводились статистическое моделирование методом Монте-Карло на основе нормального закона распределения и распределения Вейбулла-Гнеденко и бутстреп-модели.

Таблица 1

Результаты статистического моделирования нормальной случайно цензурированной выборки

Table 1

Results of statistical modeling of a normal randomly censored sample

$n=10$					$n=20$				
r	\hat{a}	$\hat{\sigma}$	$\hat{\sigma} / \sigma$	$\hat{\beta}$	r	\hat{a}	$\hat{\sigma}$	$\hat{\sigma} / \sigma$	$\hat{\beta}$
0	3,0005	0,3012	1,0039	0,9185	0	3,0026	0,2998	0,9994	0,9135
1	3,0176	0,2988	0,9960	0,9230	2	3,0251	0,3020	1,0068	0,9150
2	3,0347	0,2956	0,9852	0,9075	4	3,0497	0,3032	1,0106	0,8985
3	3,0582	0,2970	0,9899	0,9165	6	3,0715	0,3042	1,0141	0,9040
4	3,0786	0,2957	0,9857	0,9070	8	3,0932	0,3077	1,0257	0,8740
5	3,0971	0,2900	0,9667	0,8810	10	3,1134	0,3038	1,0128	0,8250

Примечание. Исходные данные: коэффициент вариации $\gamma=0,1$; параметр сдвига (математическое ожидание) $a=3,0$; параметр масштаба (среднее квадратичное отклонение) $\sigma=0,3$; уровень квантиля распределения $P=0,01$; доверительная вероятность $\beta=0,9$; \hat{a} , $\hat{\sigma}$ – МНК-оценки параметров сдвига и масштаба; n – объем выборки; r – количество цензурированных элементов выборки.

Таблица 2

Логарифмы долговечностей лопаток компрессора авиадвигателя

Table 2

Logarithms of lifetimes of aircraft engine compressor blades

4,6730*	4,7419	4,7888*	4,8215	4,8506	4,8704*	4,9111	4,9253	4,9628	4,9800
5,0607*	5,0899	5,1271	5,1523	5,1847*	5,2148	5,2430	5,2856	5,3444	5,5079*

Таблица 3

Логарифмы долговечностей лопаток компрессора авиадвигателя после моделирования цензурированных элементов выборки

Table 3

Logarithms of lifetimes of aircraft engine compressor blades after modeling censored sample elements

4,8506	4,7419	5,2856	4,8215	4,8506	4,9253	4,9111	4,9253	4,9628	4,9800
5,1271	5,0899	5,1271	5,1523	5,2430	5,2148	5,2430	5,2856	5,3444	5,5079

Таблица 4

Статистическая обработка результатов усталостных испытаний лопаток компрессора авиадвигателя

Table 4

Statistical processing of fatigue test results of aircraft engine compressor blades

$\hat{a}_{lg N}$ МНК	$\hat{\sigma}_{lg N}$ МНК	$\hat{a}_{lg N}$	$\hat{\sigma}_{lg N}$	$D\{\hat{a}\}$ МНК	$D\{\hat{\sigma}\}$ МНК	\hat{x}_{pl}	\hat{x}_p	\hat{x}_{pu}
5,0794	0,2096	5,0795	0,2058	0,0355	0,02603	4,3710	4,5918	4,7139

лирование в условиях случайного цензурирования, показавшее относительную погрешность в оценке параметров сдвига и масштаба в пределах 1–2 % и около 5 % в оценке доверительных вероятностей для квантилей распределения, что позволяет рекомендовать полученные решения для дальнейшего исследования и практического применения.

Литература

1. Cohen A.C. Progressively censored sampling in the three parameter log-normal distribution. *Technometrics*, 1976, vol. 18, no. 1, pp. 99–103.
2. Cohen A.C. Multi-censored sampling in the three para-

meter weibull distribution. *Technometrics*, 1975, vol. 17, no. 3, pp. 347–350.

3. Кендалл М.Дж., Стьюарт А. Теория распределений. М.: Наука, 1966. 588 с.
4. Кендалл М.Дж., Стьюарт А. Статистические выводы и связи. М.: Наука, 1973, 899 с.
5. Агамиров Л.В. Методы статистического анализа механических испытаний. М.: Интермет Инжиниринг, 2004. 128 с.
6. Агамиров Л.В., Агамиров В.Л., Вестяк В.А. Метод расчета квантилей распределения характеристик усталостных свойств элементов конструкций // Вестн. МАИ. 2011. Т. 18. № 4. С. 71–76.
7. Райхер В.Л. Рассеяние усталостной долговечности. М.: ЛАТМЭС, 2003. 224 с.
8. Райхер В.Л. Усталостная повреждаемость. М.: ЛАТМЭС, 2006. 238 с.

Software & Systems

DOI: 10.15827/0236-235X.117.124-129

Received 11.07.16

2017, vol. 30, no. 1, pp. 124–129

STATISTICAL ANALYSIS OF TEST RESULTS OF PRODUCTS OF AERONAUTICAL ENGINEERING IN TERMS OF RANDOM CENSORING

L.V. Agamirov¹, Dr.Sc. (Engineering), Professor, mmk@mati.ru

V.L. Agamirov², Ph.D. (Engineering), avl095@mail.ru

V.A. Vestyak², Ph.D. (Physics and Mathematics), Associate Professor, kaf311@yandex.ru,

¹ National Research University "MPEI", Krasnokazarmennaya St. 14, Moscow, 111250, Russian Federation

² Moscow Aviation Institute (National Research University), Volokolamskoe Highway, 4, Moscow, 125993, Russian Federation

Abstract. The article considers a technique of point and interval estimation of distribution parameters applied in statistical analysis of fatigue tests of aircraft structural elements based on the least squares method. The technique considers censored observations.

Relevance of the study is defined by the fact that when estimating distribution parameters for fatigue properties characteristics for a statistical analysis of fatigue tests of aircraft equipment it is necessary to take into account the results of the samples with the test finished before reaching a critical condition. The solution of this problem using known methods (maximum likelihood method) is complicated due to objective function nonmonotonicity, a number of local extremes, etc.

The first part of the article is devoted to a technique of estimating distribution parameters of observed random variables for a complete sample, which were transformed from repeatedly censored (incomplete) sample by bootstrap simulation based on order statistics. Transformation of an original randomly censored sample into a quasicomplete one is carried out in order to use the least squares method to estimate distribution parameters since this method is applicable only for a complete sample.

The second part of the article is devoted to construction of confidence limits for a quintile of observed random variable distribution. In the aircraft engineering it is applicable for assessment of a guaranteed resource normalized on a lower confidence limit of a durability quintile.

The article considers a reduction technique of repeatedly censored incomplete sample in a general case to an equivalent quasicomplete sample, for which it is possible to use the least squares method and receive the most stable and efficient evaluation with minimum dispersion. Thus, the problem of point and interval estimation of distribution parameters of fatigue properties characteristics of aircraft structure elements considering multicensored observations is solved.

Keywords: survivability, bootstrapped Modeler, least squares method, order statistics, random censoring.

References

1. Cohen A.C. Progressively censored sampling in the three parameter log-normal distribution. *Technometrics*. 1976, vol. 18, no. 1, pp. 99–103.
2. Cohen A.C. Multi-censored sampling in the three parameter weibull distribution. *Technometrics*. 1975, vol. 17, no. 3, pp. 347–350.
3. Kendall M.G., Stuart A. *The Advanced Theory of Statistics: Distribution Theory*. 1958, vol. 1, 433 p.
4. Kendall M.G., Stuart A. *The Advanced Theory of Statistics: Inference and Relationship*. 1961, vol. 2, 676 p.
5. Agamirov L.V. *Metody statisticheskogo analiza mekhanicheskikh ispytany* [The Statistical Analysis Method of Mechanical Test]. Moscow, Intermet Inzhiniring Publ., 2004, 128 p.
6. Agamirov L.V., Agamirov V.L., Vestyak V.A. A calculation method for obtaining quantiles of fatigue characteristics distribution of constructive elements. *Vestnik MAI*. 2011, vol. 18, no. 4, pp. 71–76.
7. Raykher V.L. *Rasseyanie ustalostnoy dolgozhechnosti* [Fatigue Life Scattering]. Moscow, LATMES Publ., 2003, 224 p.
8. Raykher V.L. *Ustalostnaya povrezhdaemost* [Fatigue Damaging]. Moscow, LATMES Publ., 2006, 238 p.

УДК 004.932

DOI: 10.15827/0236-235X.117.130-137

Дата подачи статьи: 16.12.16

2017. Т. 30. № 1. С. 130–137

АЛГОРИТМ ДЕТЕКТИРОВАНИЯ ОБЪЕКТОВ НА ФОТОСНИМКАХ С НИЗКИМ КАЧЕСТВОМ ИЗОБРАЖЕНИЯ

*А.С. Викторов, аспирант, alseivictor@mail.ru
(Костромской государственный университет,
ул. Дзержинского, 17, г. Кострома, 156005, Россия)*

В статье рассматривается набор алгоритмов, применяемых для распознавания объектов определенного класса на фотоснимках с некачественным изображением, полученных с видеокамеры низкого разрешения.

Особенностью рассматриваемой методики детектирования объектов является возможность обнаружения объектов, размеры изображений которых на фотоснимках не превышают нескольких десятков пикселей. Исследуемое изображение сканируется скользящим окном, считывающим участки изображения с заданным перекрытием между соседними участками. Сканируемые участки изображения предварительно обрабатываются дискриминативным автокодировщиком, извлекающим вектор признаков из участка изображения, который анализируется мультиклассовым классификатором, построенным на основе вероятностной модели регрессии, на предмет наличия изображения или части изображения объекта. Для каждого сканируемого участка изображения классификатор вычисляет значение вероятности обнаружения детектируемого объекта определенного класса на данном участке. На основании результатов сканирования изображения делается вывод о наличии изображения объекта и о его наиболее вероятном положении на фотоснимке.

Для повышения точности обнаружения границ изображения значение вероятности обнаружения детектируемого объекта определенного класса интерполируется для каждого анализируемого пикселя изображения. После детектирования пикселей на основании их распределения на изображении уточняются границы изображения детектируемого объекта.

В ходе проведенного исследования было обнаружено, что использование дискриминативного автокодировщика значительно повысило робастность алгоритма детектирования. В статье дано подробное описание процесса обучения и настройки параметров алгоритмов, используемых в процессе детектирования. Результаты данного исследования могут найти широкое применение для автоматизации различных процессов, например для сбора и анализа информации в различных аналитических системах.

Ключевые слова: каскадный шумоподавляющий автокодировщик, метод релевантных векторов, нейронная сеть, функция потерь, обучающая выборка, признаковое описание объекта, детектирование объектов, функции правдоподобия.

Основной подход, предлагаемый в современной литературе для детектирования объектов на изображении, заключается в использовании 2D-признаков, инвариантных к возможному искажению изображения объекта, вызванным аффинными преобразованиями [1], деформациями сдвига, изменением освещения или масштаба изображения объекта [2] и т.д. Для извлечения признаков из изображений, например, широко используется алгоритм SIFT (Scale-invariant feature transform), впервые рассмотренный в [3]. Для достижения качественного результата в процессе обучения классификатора при использовании признакового описания объекта путем поиска особых точек и вычисления дескрипторов их окрестностей требуется специальная подготовка обучающего набора исходных изображений и отбора уникальных признаков объекта для формирования визуального словаря, например, по методике, описанной в [4]. После извлечения 2D-признаков из тестового изображения производится их сравнение с признаками [5], хранящимися в БД. Чтобы распознать некоторый известный объект, система распознавания сначала извлекает множество характерных точек из изображений объекта, сделанных с различного ракурса, и запоминает извлеченные из них признаки в проиндексированную структуру, например та-

кую, как дерево поиска. Во время распознавания признаки извлекаются из проверяемого изображения и сравниваются с сохраненными признаками объекта. Каждый раз, когда заданное число признаков, извлеченных из тестового изображения, совпадает с признаками, описывающими детектируемый объект, система распознавания вызывает процедуру верификации – проверку на взаимное совпадение пространственного положения множества точек на тестовом изображении, из которого были извлечены признаки, с взаимным положением характерных точек, описывающих объект.

При формировании признакового описания объекта по данной методике возникает сложность процедуры формирования набора признаков, по которому однозначно можно идентифицировать объект, что приводит к появлению ложных распознаваний. Появление ложных распознаваний связано с тем, что исходные изображения определенного класса объектов могут содержать множество помех и похожие признаки могут принадлежать объектам различных классов. Поэтому для повышения надежности распознавания требуется извлечь из тренировочного набора изображений как можно больше признаков и сформировать словарь визуальных слов [6] для формирования набора уникальных признаков, что невозможно эффективно

реализовать для задачи детектирования объектов при низком разрешении изображения данных объектов. Задача поиска объектов определенного класса на фотоснимках, полученных с видеокамеры наружного наблюдения, например задача распознавания транспортных средств на этих фотоснимках, связана с детектированием объектов, как правило, имеющих низкое разрешение. Для решения данной задачи необходим иной подход к формированию признакового описания объекта.

Обзор современной литературы, описывающей методику формирования признакового описания объекта, показал, что наиболее передовой подход при формировании описания объекта заключается в использовании автокодировщика [7], являющегося нейронной сетью с симметричной архитектурой. Автокодировщик состоит из входного и выходного слоев одинаковой размерности и скрытых слоев с меньшей размерностью, чем у входного и выходного слоев. Автокодировщик имеет скрытый слой, называемый bottleneck-слоем, у которого самая наименьшая размерность из всех скрытых слоев. Размерность слоев автокодировщика при переборе от входного слоя к bottleneck-слою последовательно уменьшается, а при переборе от bottleneck-слоя к выходному слою последовательно увеличивается. Набор слоев автокодировщика от входного до bottleneck-слоя называется кодировщиком, который осуществляет преобразование некоторого входного сигнала $\bar{x}_0 \in R^n$ в сигнал $\bar{x}_{bottleneck} \in R^r$, где r – количество нейронов в bottleneck-слое. Набор слоев автокодировщика от bottleneck-слоя до выходного называется декодировщиком, он осуществляет преобразование выходного сигнала с кодировщика $\bar{x}_{bottleneck} \in R^r$ в $\bar{x}_L \in R^n$, где L – количество слоев в автокодировщике. В общем случае задача обучения автокодировщика заключается в поиске такой конфигурации весов слоев автокодировщика $\bar{W} = [\bar{W}_1, \bar{W}_2, \dots, \bar{W}_L]$ и пороговых значений $\bar{b} = [\bar{b}_1, \bar{b}_2, \dots, \bar{b}_L]$,

которая при подаче на его вход многомерного сигнала обеспечивала бы на выходе отклик, наиболее близкий к входному сигналу. В качестве функции активации для слоев автокодировщика используют нелинейную функцию активации, например сигмоид, что позволяет автокодировщику аппроксимировать сложные зависимости. Автокодировщик может применяться для уменьшения размерности входного сигнала. Уменьшение размерности входного сигнала при его подаче на вход автокодировщика осуществляется при считывании с bottleneck-слоя выходного сигнала, составляющие которого слабо коррелируют или не коррелируют между собой. Выходной сигнал с bottleneck-слоя автокодировщика далее подается на вход модуля мультиномиальной регрессии RVM (relevance vector

machine) [8], который вычисляет вероятность наличия изображения объекта определенного класса на извлеченном участке фотоснимка. По результатам сканирования всего фотоснимка производятся анализ наличия изображений объектов определенного класса на фотоснимке и определение областей фотоснимка, содержащих данные изображения.

Разработка архитектуры автокодировщика и способа его обучения

Обозначим через $\bar{x}_0 \in R^n$ входной сигнал, имеющий размерность n , а через $\bar{x}_{i+1} = f_{i+1}(\bar{x}_i) = \sigma(\bar{W}_{i+1} \cdot \bar{x}_i + \bar{b}_{i+1})$, где $i = 0, \dots, L-1$, функциональную зависимость между сигналами на входе \bar{x}_i и выходе \bar{x}_{i+1} $i+1$ -го слоя, где \bar{W}_{i+1} – весовые коэффициенты $i+1$ -го слоя; \bar{b}_{i+1} – пороговые значения $i+1$ -го слоя; $\sigma(\bar{x}) = \frac{1}{1 + e^{-\bar{x}}}$ – функция ак-

тивации, в качестве которой согласно [9] выбрана сигмоид-функция. Далее запишем функциональную зависимость между входом и выходом автокодировщика в виде $\bar{x}_L = DF(\bar{x}_{bottleneck}) = DF(EF(\bar{x}_0))$,

где $EF(\bar{x}_0) = f_k(f_{k-1}(\dots(f_1(\bar{x}_0))))$ – функциональная зависимость между сигналами на входе и выходе кодировщика; $DF(\bar{x}_{bottleneck}) = f_L(f_{L-1}(\dots(f_{k+1}(\bar{x}_{bottleneck}))))$ – функциональная зависимость между сигналами на входе и выходе декодировщика; L – число слоев автокодировщика.

Задача настройки автокодировщика заключается в поиске такой конфигурации $C(\bar{W}, \bar{b})$ весов \bar{W} и пороговых значений \bar{b} , которая минимизировала бы выбранную целевую функцию. В качестве целевой функции используется функция потерь $L(\bar{x}_0, \bar{x}_L)$, определяющая меру несоответствия сигнала на выходе автокодировщика от ожидаемого сигнала при подаче на его вход заданного сигнала. Для стандартного автокодировщика, предназначенного только для сжатия информации, функция потерь имеет вид

$$L(\bar{x}_0, \bar{x}_L) = \|\bar{x}_0 - \bar{x}_L\|_2^2 = \|\bar{x}_0 - DF(EF(\bar{x}_0))\|_2^2,$$

а задача обучения автокодировщика [10] сводится к минимизации целевой функции вида

$$C(\bar{W}, \bar{b}) = \arg \min_{\bar{W}, \bar{b}} \frac{1}{2} \cdot \sum_{i=0}^{N-1} \|\bar{x}_{0,i} - DF(EF(\bar{x}_{0,i}))\|_2^2 + \frac{1}{2} \cdot \lambda \cdot \|\bar{W}\|_F^2, \text{ где } N - \text{размер обучающей выборки } X.$$

Для обучения автокодировщика дополнительному функционалу, а именно способности извлекать информацию о заданной категории объектов, необходимо использовать другую функцию потерь, которая учитывала бы наличие объекта, относящегося к определенной категории на изображении.

Для обучения автокодировщика, обладающего описанным функционалом, имеется обучающий набор изображений $X \in \chi^+ \cup \chi^-$, состоящий из N изображений. Каждому изображению из набора назначена метка класса $l(\bar{x}_0) \in \{-1, 1\}$. Каждое изображение из набора имеет размер $m \times m$, при этом $m^2 = n$, где n – размерность входа автокодировщика. Изображение из набора имеет метку 1, если содержит изображение всего или части объекта, принадлежащего к классу детектируемых объектов, в противном случае изображение имеет метку -1. В качестве функции потерь для выбранного автокодировщика предлагается модифицировать стандартную целевую функцию путем добавления в нее дополнительной функции потерь, например, как в работе [11], для учета принадлежности входного сигнала, подаваемого на вход автокодировщика к детектируемому классу объектов.

Обозначим через $\bar{z}^t = \{\bar{z}_{-1}^t, \bar{z}_1^t\}$ набор состояний bottleneck-слоя автокодировщика при его обучении, где $\bar{z}_l^t = \{\bar{z}_{l,1}^t, \dots, \bar{z}_{l,j}^t, \dots, \bar{z}_{l,J}^t\}$ – состояние скрытого k -го bottleneck-слоя автокодировщика при подаче на его вход изображений из обучающего набора, имеющих определенную метку из набора $l \in \{-1, 1\}$, где $j=1, \dots, J$ – номер образца изображения из поднабора изображений, имеющих определенную метку; t – номер итерации процесса обучения. Тогда дополнительную функцию потерь можно записать в виде

$$L(\bar{z}^t) = \sum_{x \in \chi^+ \cup \chi^-} \max \left(0, l(x) \cdot \left(\frac{1}{|\chi^{l(x)}| - 1} \times \right. \right. \quad (1)$$

$$\left. \left. \times \sum_{\forall k | k \neq j} \left\| \bar{z}_{l(x), j}^t - \bar{z}_{l(x), k}^t \right\| - \log \frac{1}{|\chi^{-l(x)}|} \sum_{\forall k} e^{\left(\left\| \bar{z}_{l(x), j}^t - \bar{z}_{-l(x), k}^t \right\| \right)} - thr \right) \right),$$

где $|\chi^{l(x)}|$ – мощность подмножества образцов из обучающего набора с заданной меткой; \log – обозначение десятичного логарифма; $\left\| \bar{z}_{l(x), j}^t - \bar{z}_{l(x), k}^t \right\|$ – евклидово расстояние между состояниями bottleneck-слоя автокодировщика (при подаче на вход двух различных изображений); thr – пороговое значение, задающее диапазон разброса расстояний между состояниями bottleneck-слоя автокодировщика.

Так как на снимках, полученных с видеокамеры наружного наблюдения, размер изображений объектов, как правило, не превышает 100×100 пикселей, для их эффективного поиска методом скользящего окна экспериментально выбран размер окна 20×20 . Для обнаружения объекта изображение сканируется скользящим окном, которое последовательно перемещается по заданной направляющей сетке, обеспечивающей перекрытие между соседними сканируемыми участками изображения в 40 %. Так как окно имеет размер 20×20 пикселей, перекрытие составляет 8 пикселей. Полученные при помощи аппаратуры видеонаблюдения изображения могут содержать сильную шумовую компоненту, например из-за плохой видимости или внешнего освещения, поэтому для качественной реализации решения задачи классификации необходимо производить предварительную фильтрацию изображений.

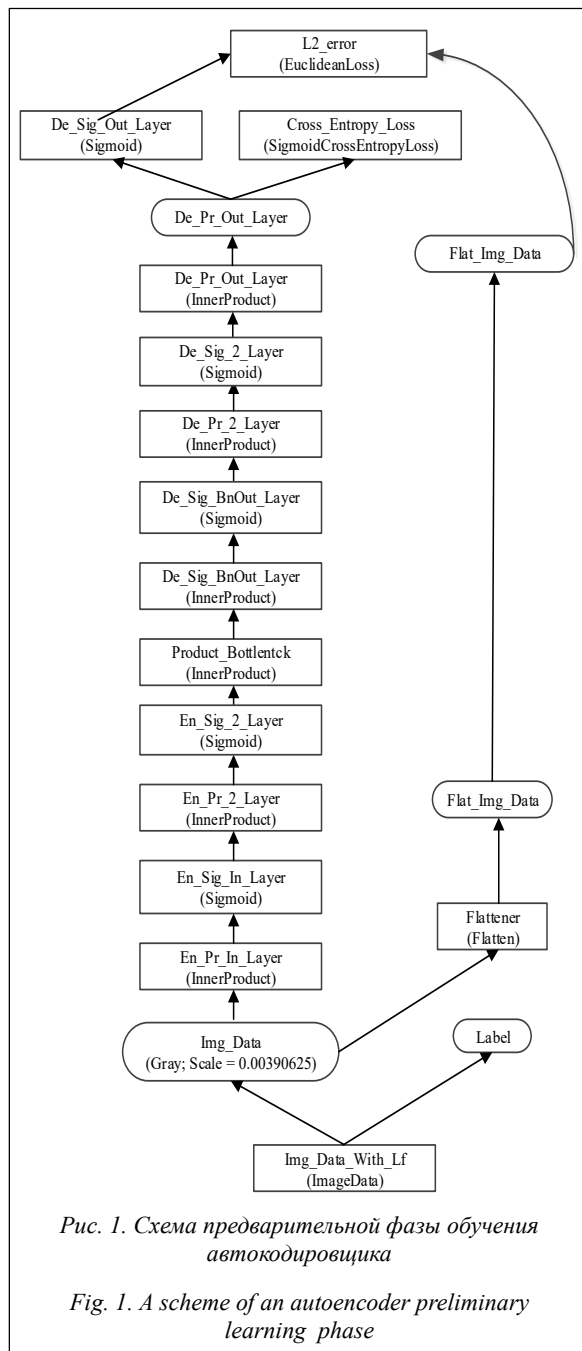
В результате поиска возможных архитектурных решений для реализации автокодировщика и методик обучения автокодировщика была выбрана концепция автокодировщика SDA (stacked denoising autoencoder) [12], который позволяет производить не только уменьшение размерности входных данных, но и их фильтрацию от возможных помех. За основу реализации процесса обучения автокодировщика была выбрана методика процесса обучения, описанная в статье [13].

Для обучения автокодировщика разработано ПО, для реализации которого используется фреймворк Caffe, который был разработан в BLVC (Berkeley Vision and Learning Center). Обучение производится в две стадии. На первой стадии автокодировщик учится восстанавливать образ, поданный на его вход, как стандартный автокодировщик, по схеме, изображенной на рисунке 1.

Для обучения автокодировщика во время проведения первой фазы обучения из обучающего набора используются только изображения, относящиеся к классу детектируемых объектов (то есть имеющие метку 1). После завершения первой фазы производится обучение обученного стандартному функционалу автокодировщика дополнительному функционалу с использованием функции потерь (1), которая производит анализ сигнала на выходе bottleneck-слоя, и функции потерь из [11], которая производит анализ сигнала выходного слоя автокодировщика (на рисунке 1 слой с названием De_Sig_Out_Layer).

Для обучения автокодировщика цветное изображение предварительно преобразуется в черно-белое, после чего производится трансформация значений интенсивности пикселей изображения из формата $[0, 255]$ в $[0, 1]$ путем умножения значений на масштабирующий коэффициент $sk = 0,00390625$.

После обучения автокодировщика производится обучение классификатора RVM, с которым



автокодировщик будет работать совместно, в соответствии со схемой, изображенной на рисунке 2.

После подачи на вход автокодировщика участка исходного изображения с его bottleneck-слоя считывается многомерный сигнал $\bar{x}_{bottleneck} = EF(\bar{x}_{0, new}) \in R^{30}$, который является сжатым и отфильтрованным образом участка исходного изображения. Полученный сигнал подается на вход классификатора RVM, обучение которого производится на маркированном наборе образов, считанных с bottleneck-слоя обученного автокодировщика при подаче на его вход изображений из обучающего набора. Изображения, используемые

для обучения автокодировщика, могут быть искусственно синтезированы из исходных изображений обучающего набора путем искусственных трансформаций и добавления шумовой составляющей.

Метод релевантных векторов для задачи распознавания

Пусть $(\bar{t}, \bar{x}) = \{t_n, \bar{x}_n\}_{n=1}^N$ – обучающая выборка, полученная из обучающего набора изображений после их обработки автокодировщиком, где $\bar{x}_n = x_{bottleneck, n} \in R^r$ – вектор признаков, извлеченный из изображения обучающего набора при помощи автокодировщика; N – число изображений в обучающем наборе; $\bar{t}_n = (t_0, \dots, t_K) \in R^K$ – вектор, j -й элемент которого равен 1, а остальные элементы нулевые, что указывает на принадлежность рассматриваемого образца к j -му классу объектов, при этом метки от 1 до K соответствуют различным классам детектируемых объектов (например, класс детектируемых объектов «автотранспортные средства» состоит из подклассов «легковые», «грузовые» средства и т.д.).

Необходимо определить, содержит ли изображение, представленное своим вектором признаков $\bar{x} \in R^r$, извлеченным при помощи автокодировщика, изображение объектов определенного класса; если да, то вывести метку класса изображенного объекта.

Рассмотрим сущность метода RVM, предназначенного для решения задачи восстановления регрессии, при этом поиск решения осуществляется в классе линейных решающих правил [14]:

$$y = f(\bar{x}, \bar{w}) + \varepsilon = \varepsilon + \sum_{i=0}^M w_i \cdot \phi_i(\bar{x}) = \bar{w}^T \cdot \bar{\phi}(\bar{x}) + \varepsilon,$$

где \bar{x} – вектор признаков; y – скалярное значение;

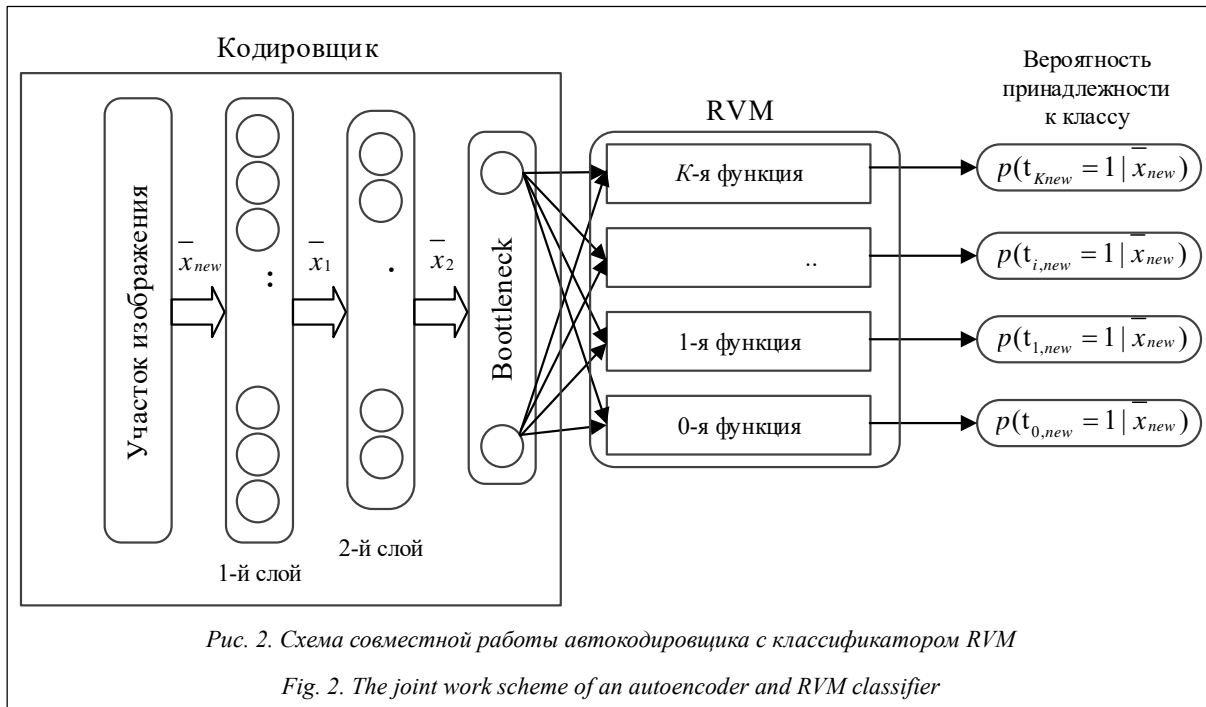
$\bar{\phi}(\bar{x}) = (\phi_0(\bar{x}), \dots, \phi_M(\bar{x}))^T$ – набор базисных функций (в качестве базисной функции в данной реализации алгоритма используется гауссова радиальная

базисная функция); $\bar{w} = (w_0, \dots, w_M)^T$ – весовые коэффициенты, которые определяют вклад, вносимый определенной базисной функцией;

$\varepsilon \sim N(0, \sigma^2)$ – аддитивный гауссовский шум с нулевым математическим ожиданием. Для обучения рассматриваемой модели используется метод максимального правдоподобия, который позволяет производить оценку весовых коэффициентов по обучающей выборке:

$$\bar{w}_{est} = \arg \min_{\bar{w}} \left(\left\| \bar{y} - \bar{\Phi} \cdot \bar{w} \right\|^2 \right),$$

где $\bar{y} = (y_1, \dots, y_N)^T$ – набор скалярных значений, каждое из которых соответствует определенному вектору \bar{x}_n из обучающей выборки $\bar{x} = (\bar{x}_1, \dots, \bar{x}_N)^T$;



$\bar{\Phi} = (\bar{\phi}(\bar{x}_1), \dots, \bar{\phi}(\bar{x}_N))^T$ – матрица, сформированная из значений базисных функций, вычисленных для обучающей выборки. Для адаптации RVM к задаче бинарной классификации в статье [14] рассматривается функция правдоподобия, основанная на распределении Бернулли:

$$p(\bar{t} | \bar{w}) = \prod_{i=1}^N \sigma \left\{ \left(y(\bar{x}_i) \right) \right\}^{t_i} \cdot \left[1 - \sigma \left\{ \left(y(\bar{x}_i) \right) \right\} \right]^{1-t_i}, \quad (2)$$

где $y(\bar{x}) = \sigma(f(\bar{x}, \bar{w})) = \frac{1}{1 + \exp(-\bar{w}^T \cdot \bar{\phi}(\bar{x}))}$ – логистическая функция; $(\bar{t}, \bar{x}) = \{t_n, \bar{x}_n\}_{n=1}^N$ – обучающая выборка; $t_n = t(\bar{x}_n) \in \{0, 1\}$ – элементы вектора \bar{t} .

Стандартный подход к решению задачи мультиклассовой классификации, рассматриваемой в данной статье, – мультиномиальная логистическая регрессия $p(t_{j,new} = 1 | \bar{x}_{new}) = \frac{\exp(y_j(\bar{x}_{new}))}{\sum_{k=1}^K \exp(y_k(\bar{x}_{new}))}$, где

примем $y(\bar{x}) = \bar{w}^T \cdot \bar{\phi}(\bar{x})$; $j=0, \dots, K$. Тогда аналогичным выражению (2) образом вводим следующую функцию правдоподобия:

$$p(\bar{t} | \bar{w}) = \prod_{i=1}^N \left(\frac{\exp(\bar{w}_{til}^T \cdot \bar{\phi}_{til}(\bar{x}_i))}{\sum_{k=0}^K \exp(\bar{w}_k^T \cdot \bar{\phi}_k(\bar{x}_i))} \right),$$

где til – метка класса i -го образца. Для оценки апостериорного максимума выведенной функции

правдоподобия используем технику, описанную в статье [14]. При этом делаем допущение о распределении весовых коэффициентов $p(\bar{w} | \bar{\alpha}) = \prod_{i=0}^M N(w_i | 0, \alpha_i^{-1})$, где $\bar{\alpha} = (\alpha_0, \dots, \alpha_M)^T$ – вектор гиперпараметров, и шума измерения $\bar{\varepsilon} \sim N(0, \beta^{-1} \cdot \bar{I})$, а также вводим эквивалентную объектную функцию $J(\bar{w}) = \sum_{i=1}^N \log p(t_i | \bar{w}_{til}) + \sum_{i=1}^N \log p(\bar{w}_{til} | \alpha_{til}^*)$ [14].

Реализация процесса сканирования изображения

Процесс детектирования объектов определенного класса разбит на несколько фаз. Во время первой фазы фотоснимок сканируется скользящим окном, которое перемещается по регулярной сетке, обеспечивая 40 %-ное перекрытие между соседними сканируемыми участками фотоснимка.

После первой, предварительной, фазы детектирования производится отбор участков фотоснимка, для которых предсказанное RVM-значение вероятности $p(t_k = 1 | \bar{x})$ того, что данный участок фотоснимка содержит изображение детектируемого объекта определенного класса k , больше некоторого заданного порогового p_{thrMax} значения. Для каждого отобранного участка вычисляется положение центрального пикселя, которому назначается метка класса k , выявленная модулем RVM. Далее для всех пикселей, соседних с центральным пикселем, вычисляется принадлежность к классу объектов с меткой k . Вычисление принадлежности рас-

смаатриваемых пикселей к данному классу k производится на основании значения вероятности, которое вычисляется путем интерполяции методом natural neighbor [15]. Интерполяция значений вероятности $p(t_k = 1 | \bar{x})$ принадлежности к данному классу объектов k производится на основе данных о значении $p(t_k = 1 | \bar{x})$ для центральных участков пикселей, соседних с рассматриваемым участком. Если значение $p(t_k = 1 | \bar{x})$ для пикселя, принадлежность которого к классу объектов k выясняется, больше некоторого заданного порогового p_{thrMin} значения, данному пикселю назначается метка класса объектов k . Если выясняется, что исследуемый пиксель принадлежит к классу объектов k , производится выявление принадлежности или непринадлежности к классу объектов k соседних с ним пикселей, принадлежность которых еще не выявлена. В противном случае, если исследуемый пиксель не принадлежит к классу объектов k , процедура выявления принадлежности соседних с ним пикселей не производится.

В результате описанной процедуры формируется множество пикселей, для которых вычисляется минимальный ограничивающий прямоугольник, определяющий границы изображения детектируемого объекта.

Экспериментальная часть

Для оценки качества детектирования объектов, имеющих низкое разрешение, и целесообразности использования разработанного алгоритма был проведен эксперимент, в котором для обучения детектора и оценки качества его работы использовался полученный с веб-ресурса набор изображений аэрофотосъемки HRO_2012_6_Inch_Orthophotography [16]. Изображения из набора имеют разрешение 12,5×12,5 см на пиксель.

Из данного набора были сформированы два набора изображений, один из которых содержит изображения транспортных средств, имеющих размеры примерно 15×35 пикселей. Второй набор изображений, полученный из исходного набора, содержит различные изображения земной поверхности без изображений транспортных средств.

Обучение детектора и оценка качества детектирования производились на персональном компьютере, обладающем следующими характеристиками: центральный процессор – Intel Core i7-6700K; оперативная память – 8 Гб; видеопроцессор – NVIDIA GEFORCE GTX 980 Ti; операционная система – Ubuntu 14.04 LTS. Для измерения качества детектирования использовались следующие показатели: MAP (mean average precision); 0.01 FPPI (false positive per image); 0.1 FPPI; 1 FPPI. Сравнение качества детектирования разработан-

ного алгоритма производилось с DPM (deformable part model) [17], детектором HOG+SVM [18], детектором RVM, работающим в связке со стандартным автокодировщиком.

Результаты оценки работы алгоритмов приведены в таблице.

Результаты эксперимента

Experiment results

Детектор	MAP	0.01 FPPI	0.1 FPPI	1 FPPI
DPM	62.7±4.3	14.4±7.1	34.5±6.1	76.5±5.2
HOG+SVM	58.9±4.0	12.9±6.2	32.1±4.8	74.8±4.7
Standard Autoencoder+RVM	59.8±3.9	13.9±5.6	32.9±4.2	75.1±4.6
Discriminative Autoencoder+RVM	79.8±4.2	31.5±5.2	58.0±4.6	89.5±2.5

Проведенный эксперимент показал, что RVM в связке со стандартным автокодировщиком работает хуже DPM, но лучше детектора HOG+SVM. Если вместо стандартного автокодировщика использовать дискриминативный автокодировщик, качество детектирования значительно возрастет, то есть данный алгоритм превзойдет остальные алгоритмы, включая алгоритм DPM. Таким образом, результаты исследования доказывают целесообразность использования предложенного алгоритма.

Выводы

В результате проведенного исследования был предложен алгоритм детектирования изображений объектов определенного класса на фотоснимках, полученных с видеокамеры при малом разрешении изображений детектируемых объектов и низком качестве фотоснимков. В процессе создания алгоритма выбрана архитектура и разработана схема обучения автокодировщика, предназначенного для формирования признаков описания объектов по их изображениям.

В работе предложена оригинальная функция потерь для обучения автокодировщика дополнительному функционалу, а именно способности извлекать информацию только о заданной категории объектов.

Для детектирования объектов использована схема скользящего окна, которое сканирует фотоснимок, обеспечивая перекрытие между соседними сканируемыми участками фотоснимка в 40 %. С целью повышения точности детектирования границ изображения объекта применен мультиномиальный классификатор, построенный на основе алгоритма RVM, для вычисления вероятности нахождения изображения объекта или его частей в сканируемом участке фотоснимка. Также для по-

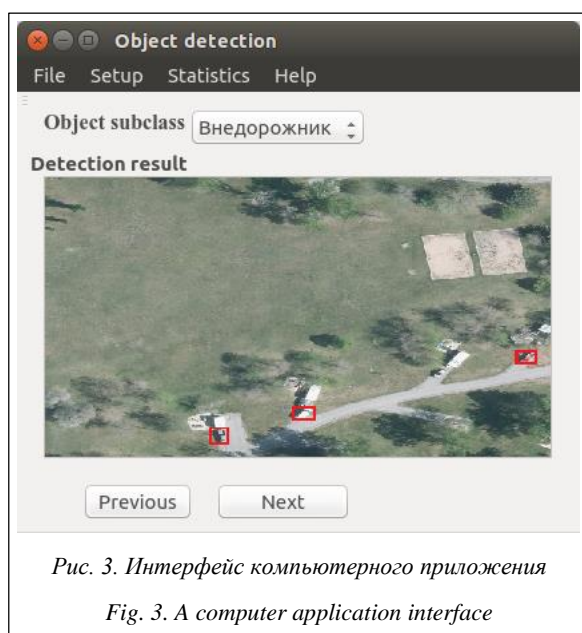


Рис. 3. Интерфейс компьютерного приложения

Fig. 3. A computer application interface

вышения точности обнаружения границ изображения детектируемого объекта было предложено интерполировать значения вероятности обнаружения детектируемого объекта определенного класса для каждого пикселя изображения, анализ принадлежности которого к изображению объекта проводится. После чего на основании распределения пикселей, которые были приняты за принадлежащие к изображению объекта определенного класса, производится оценка границ изображения детектируемого объекта. На основании результатов проведенного исследования было разработано компьютерное приложение (рис. 3).

Результаты данного исследования могут найти широкое применение для автоматизации различных процессов, например, для сбора и анализа информации в различных аналитических системах.

Литература

1. Mikolajczyk K., Schmid C. An affine invariant interest point detector. Proc. 7th ECCV'02, 2002, part I, pp. 128–142.
2. Harris C.G., Stephens M. A combined corner and edge detector. Proc. of 4th Alvey Vision Conf. 1988, pp. 147–151.

3. Lowe D.G. Distinctive image features from scale-invariant keypoints. Intern. Jour., of Comp. Vision. 2004, vol. 60, iss. 2, pp. 91–110.
4. Sivic J., Zisserman A. Video Google: a text retrieval approach to object matching in videos. Proc. ICCV'03. 2003, vol. 2, pp. 1470–1477.
5. Schaffalitzky F., Zisserman A. Automated location matching in movies. Comp. Vision and Image Understanding. 2003, vol. 92, pp. 236–264.
6. Fergus R., Fei-Fei L., Perona P., Zisserman A. Learning object categories from Google's image search. Proc. Intern. Conf. on Comp. Vision. 2005, vol. 2, pp. 1816–1823.
7. Krizhevsky A., Hinton G.E. Using very deep autoencoders for content-based image retrieval. Conf. ESANN 2011, 19th Europ. Sympos. on Artificial Neural Networks, Bruges, Belgium, 2011, pp. 44–51.
8. Tipping M.E. Sparse bayesian learning and the relevance vector machine. The Jour. of Machine Learning Research. 2001, vol. 1, pp. 211–244.
9. Kamyschanska H., Memisevic R. On autoencoder scoring. ICML (3), vol. 28 of JMLR Workshop and Conf. Proc. 2013, pp. 720–728.
10. Wenchao Yu, Guangxiang Zeng, Ping Luo, Fuzhen Zhuang, Qing He, Zhongzhi Shi. Embedding with autoencoder regularization. Machine Learning and Knowledge Discovery in Databases Europ. Conf., ECML PKDD 2013, Prague, Czech Republic, 2013, Proc., part III. Berlin Heidelberg: Springer, 2013, pp. 208–223.
11. Razakarivony S., Discriminative F.J. Autoencoders for small targets detection. ICPR '14 Proc. 22nd Intern. Conf. on Pattern Recognition. 2014, pp. 3528–3533.
12. Yoonseop Kang, Kang-Tae Lee, Jihyun Eun, Sung Eun Park, Seungjin Choi. Stacked denoising autoencoders for face pose normalization. Neural Information Processing 20th Intern. Conf., ICONIP 2013, Daegu, Korea, Proc., part III. Berlin Heidelberg: Springer, 2013, pp. 241–248.
13. Kin Gwn Lore, Adedotun Akintayo, Soumik Sarkar. LLNet: a deep autoencoder approach to natural low-light image enhancement. Cornell Univ. Library. URL: <https://arxiv.org/abs/1511.03995> (дата обращения: 05.2016).
14. Tzikas D.G., Liyang Wei, Likas A., Yongyi Yang, Galatsanos N.P. A tutorial on relevance vector machines for regression and classification with applications. EURASIP Newsletter, 2012, vol. 17, no. 2, pp. 4–23.
15. Enriquez-Cervantes C.J., Rodriguez-Dagnino R.M. A super-resolution image reconstruction using natural neighbor interpolation. Computación y Sistemas, 2015, vol. 19, no. 2, pp. 211–231.
16. UTAH AGRC. URL: <https://gis.utah.gov/data/aerial-photography/> (дата обращения: 01.2016).
17. Felzenszwalb P.F., Girshick R.B., McAllester D., Ramanan D. Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, vol. 32, iss. 9, pp. 1627–1645.
18. Elhoseiny M., Bakry A., Elgammal A. MultiClass object classification in video surveillance systems – experimental study. Proc. 2013 IEEE CVPRW '13, 2013, pp. 788–793.

OBJECT DETECTION ALGORITHM IN LOW IMAGE QUALITY PHOTOGRAPHS

A.S. Viktorov¹, Postgraduate Student, alsevictor@mail.ru

¹ Kostroma State University, Dzerzhinskogo St. 17, Kostroma, 156005, Russian Federation

Abstract. The article considers a set of algorithms for specified class object recognition in low quality photographs obtained via camera with low resolution.

A special feature of the considered method of object detection is the ability to detect objects even if their sizes in images don't exceed several tens of pixels. Each processed image is scanned via sliding window of fixed width and height that reads rectangular image regions with specified overlap between neighboring regions. All scanned image regions are preliminarily processed by a discriminative autoencoder to extract feature vector from a processed image region. Further analysis of an extracted vector includes classifier means on the basis of probabilistic multinomial regression model to check the scanned region of image if there is object image or its parts. The classifier calculates the probability of detection of a certain class detectable object in each scanned image region. On the basis of an image scan result there is a conclusion on the object image presence and its most probable position in the photograph.

To improve the accuracy of calculation of detected object image boundaries, the value of a detection probability of a certain detectable object is interpolated for each pixel, which is analyzed for belonging to the image of the object. After that, on the basis of the detected pixel distribution on the image it is possible to estimate the boundaries of the detected object.

The experiment has revealed that using a discriminative autoencoder significantly increases detection algorithm robustness. The article also gives a detailed description of a learning and algorithm parameters adjustment process. The results of this research can be widely used to automate various processes, for example, to collect and analyze information in various analytical systems.

Keywords: stacked denoising autoencoders, relevance vector machine, neural net, loss function, training data set, feature vector, object detection, likelihood function.

References

1. Mikolajczyk K., Schmid C. An affine invariant interest point detector. *Proc. 7th European Conf. on Computer Vision (ECCV '02)*. Part 1, 2002, pp. 128–142.
2. Harris C.G., Stephens M. A combined corner and edge detector. *Proc. 4th Alvey Vision Conf.* 1988, pp. 147–151.
3. Lowe D.G. Distinctive image features from scale-invariant keypoints. *Int. Jour. of Computer Vision*. 2004, vol. 60, iss. 2, pp. 91–110.
4. Sivic J., Zisserman A. Video Google: a text retrieval approach to object matching in videos. *Proc. of 9th IEEE Int. Conf. on Computer Vision (ICCV '03)*. 2003, vol. 2, pp. 1470–1477.
5. Schaffalitzky F., Zisserman A. Automated location matching in movies. *Computer Vision and Image Understanding. Special iss. on video retrieval and summarization*. 2003, vol. 92, pp. 236–264.
6. Fergus R., Fei-Fei L., Perona P., Zisserman A. Learning object categories from google's image search. *Proc. Int. Conf. on Computer Vision*. 2005, vol. 2, pp. 1816–1823.
7. Krizhevsky G.E. Hinton using very deep autoencoders for content-based image retrieval. *Proc. Conf. ESANN 2011, 19th European Symp. on Artificial Neural Networks*. 2011, Bruges, Belgium, pp. 44–51.
8. Tipping M.E. Sparse bayesian learning and the relevance vector machine. *Jour. of Machine Learning Research*. 2001, vol. 1, pp. 211–244.
9. Kamyshanska H., Memisevic R. On autoencoder scoring. *Proc. ICML (3), vol. 28 of JMLR Workshop and Conf.* 2013, pp. 720–728.
10. Wenchao Yu, Guangxiang Zeng, Ping Luo, Fuzhen Zhuang, Qing He, Zhongzhi Shi. *Proc. Embedding with Autoencoder Regularization. Machine Learning and Knowledge Discovery in Databases European Conf. (ECML PKDD 2013)*, Berlin Heidelberg, Springer Publ., 2013, part III, pp. 208–223.
11. Razakarivony S., Discriminative F.J. Autoencoders for small targets detection. *Proc. 22nd Int. Conf. on Pattern Recognition (ICPR '14)*. 2014, pp. 3528–3533.
12. Kang Y.-S., Lee K.-T., Eun J.-H., Park S.-E., Choi S.-J. Stacked denoising autoencoders for face pose normalization. *Proc. Neural Information Processing 20th Int. Conf. (ICONIP 2013)*, Berlin Heidelberg, Springer Publ., 2013, part III, pp. 241–248.
13. Kin Gwn Lore, Adedotun Akintayo, Soumik Sarkar. *LLNet: a deep autoencoder approach to natural low-light image enhancement*. Cornell Univ. Library. Available at: <https://arxiv.org/abs/1511.03995> (accessed May 5, 2016).
14. Tzikas D.G., Wei L., Likas A., Yang Y., Galatsanos N.P. *A tutorial on relevance vector machines for regression and classification with applications*. Dept. of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, USA, 2012, no. 6, pp. 4–23.
15. Enriquez-Cervantes C.J., Rodriguez-Dagnino R.M. *A super-resolution image reconstruction using natural neighbor interpolation*. Monterrey, Mexico, Tecnologico de Monterrey, Electrical and Computing Engineering Dept., 2015, pp. 211–231.
16. UTAH AGRC. Available at: <https://gis.utah.gov/data/aerial-photography/> (accessed January 16, 2016).
17. Felzenszwalb P.F., Girshick R.B., McAllester D., Ramanan D. Object detection with discriminatively trained part-based models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*. 2010, vol. 32, iss. 9, pp. 1627–1645.
18. Elhoseiny M., Bakry A., Elgammal A. MultiClass Object classification in video surveillance systems – experimental study. *Proc. 2013 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW '13)*. 2013, pp. 788–793.

УДК 20.19.27

DOI: 10.15827/0236-235X.117.138-142

Дата подачи статьи: 23.11.16

2017. Т. 30. № 1. С. 138–142

АВТОМАТИЧЕСКИЙ СИНТАКСИЧЕСКИЙ АНАЛИЗ КИТАЙСКИХ ПРЕДЛОЖЕНИЙ ПРИ ОГРАНИЧЕННОМ СЛОВАРЕ

Юй Чуцяо, аспирант, *yuchiqiao123@gmail.com*;

И.А. Бессмертный, д.т.н., профессор, *bia@cs.ifmo.ru*
(Санкт-Петербургский национальный исследовательский университет
информационных технологий, механики и оптики (Университет ИТМО),
Кронверкский просп., 49, г. Санкт-Петербург, 197101, Россия)

В работе обсуждается проблема автоматического анализа естественно-языковых текстов на китайском языке. Одной из актуальных задач в этой области является автоматическое извлечение из текстовых документов фактов по запросу, поскольку автоматические переводчики здесь практически бесполезны.

Целью работы является прямое извлечение фактов из текстов на языке оригинала без его перевода. Для этого предлагается подход на основе синтаксического анализа предложений анализируемого текста с последующим сопоставлением найденных частей речи с формализованным запросом в форме субъект–предикат–объект.

Отличительная особенность предложенного алгоритма синтаксического анализа – отсутствие фазы сегментации последовательности иероглифов, составляющих предложения, на слова. Узким местом при решении данной задачи является словарь, поскольку при отсутствии слова в словаре правильная интерпретация фразы может быть невозможна. Для преодоления этой проблемы в работе предлагается идентификация модели предложения по служебным словам, а ограниченность словаря устраняется предварительным автоматическим построением тезауруса предметной области и словаря общеупотребительных слов на основе статистической обработки корпуса документов.

Апробация предложенного подхода выполнена на небольшой предметной области и с ограниченным словарем, где данный метод показал свою работоспособность. Проведен также анализ временных характеристик разработанного алгоритма.

Поскольку для синтаксического анализа используется метод простого перебора, скорость работы парсера на реальных задачах может оказаться неприемлемо низкой, что должно стать темой дальнейших исследований.

Ключевые слова: *естественные языки, синтаксический анализ, извлечение фактов, тезаурус, дерево поиска.*

В последнее время наблюдается рост потребности в поиске информации в китайских текстах. В частности, в силу расширения кооперационных связей между нашими странами возникает необходимость мониторинга сайтов госзакупок, научных фондов, производителей товаров и услуг. Существующие программы автоматического перевода, например Google Translate, здесь не очень полезны, поскольку дают только перевод, который еще нужно интерпретировать, чтобы извлечь нужную информацию. В связи с этим целесообразно прямое извлечение фактов из исходного текста на языке оригинала. Поставленная цель достигается решением следующих основных задач. Во-первых, необходимо выполнить синтаксический анализ естественно-языкового текста. Во-вторых, найти предложение в тексте, содержащее интересующую нас информацию. В-третьих, провести унификацию – сопоставить утверждения запроса с частями речи предложения. Результатом унификации как раз и будет подстановка в переменные запроса искомого значений.

Китайский язык, несмотря на кажущуюся сложность, имеет чрезвычайно простую грамматику. Главная проблема при анализе китайских текстов обусловлена двумя особенностями. В китайском языке отсутствуют пробелы между словами, и практически любое сочетание иероглифов может быть интерпретировано тем или иным образом. Это порождает проблему сегментации предложе-

ний на слова. Даже наличие полного словаря не гарантирует правильную интерпретацию последовательностей символов [1]. Статистические методы, в частности метод взаимной информации [2], позволяют обойтись без словарей, но также не дают 100 %-ную полноту и точность, поскольку в подавляющем большинстве случаев сегментация определяется контекстом, то есть в процессе семантического анализа.

Другой проблемой интерпретации китайских текстов является словарь. В отличие от алфавитных языков, где одна буква ничего не означает, в иероглифических языках каждому иероглифу соответствует определенное, достаточно широкое понятие. Конкретный смысл иероглиф приобретает только в контексте, из чего следует, что синтаксический анализ в китайском языке неотделим от семантического.

Постановка задачи

В целях извлечения фактов задача синтаксического анализа китайских текстов может быть конкретизирована следующим образом. Поисковый запрос должен быть формализован в виде *субъект–предикат–объект*, где *субъект*, *предикат* и *объект* могут быть либо словами, либо переменными. Для исполнения запроса необходимо найти в тексте фразу с заданными словами в качестве указанных членов предложений. Структура предложения

описывается его грамматической моделью. Следовательно, для каждого предложения необходимо найти его модель и подставить в него слова из поискового запроса. Полный синтаксический анализ в таком случае делать не требуется.

Состояние проблемы

Существует большое количество инструментальных средств синтаксического анализа, включая инструменты с открытыми кодами, например Томита-парсер [3], но все они ориентированы на алфавитные языки. Среди синтаксических анализаторов (парсеров), поддерживающих китайский язык, следует выделить Stanford CoreNLP [4] и SKIP Chinese Parser [5]. Основной проблемой всех парсеров, в том числе перечисленных, является структурная неоднозначность синтаксического разбора, обусловленная возможным наличием многих ролей слов в предложении. Для сокращения числа вариантов интерпретации предложений применяются методы машинного обучения, в частности [6], где учитываются частоты совместной встречаемости пар слов, а также принцип «разделяй и властвуй» (divide and conquer), при котором фраза разбивается на последовательности, анализируемые раздельно [7]. При этом качество синтаксического разбора существенно зависит от словаря. Таким образом, синтаксический анализ китайских текстов является сложной задачей, приемлемого решения которой на данный момент не существует.

Предлагаемый подход к синтаксическому анализу китайских текстов

Синтаксический анализ опирается на грамматику и словарь, однако их недостатки ухудшают качество парсинга. Если отсутствие адекватной структурной модели приводит к искаженной интерпретации фразы, то отсутствие хотя бы одного слова в словаре может сделать синтаксический анализ невозможным. Во флективных языках, в частности русском, имеется большая избыточность за счет окончаний, суффиксов, предлогов, приставок и др., что позволяет в принципе вообще обходиться без словаря [8]. К сожалению, китайский язык не обладает свойством флективности, более того, в нем отсутствуют даже времена глаголов, а одно и то же слово может использоваться в качестве глагола, существительного, наречия или прилагательного.

В работе предлагается не подключать к парсеру мощные словари, а наоборот, ограничить словарь небольшим количеством слов, в наибольшей степени определяющих структуру фразы. К таким словам относятся модальные глаголы, предлоги, послелогии, такие как 吗 (*ma* – аналог частицы *ли* в русском языке – признак вопроса), 的 (*de* – признак

притяжательного прилагательного или аналога родительного падежа), 了 и 过 (*le* и *guo* – две разновидности прошедшего времени), счетные слова и некоторые другие.

С учетом жесткого порядка слов в китайском языке выделение в предложении служебных слов с определенной достоверностью может позволить выявить его структуру. Однако этого недостаточно без сегментации последовательностей иероглифов на слова. Здесь можно использовать вышеупомянутые статистические методы [1, 2], позволяющие фиксировать границы слов на редко встречающихся парах иероглифов. В работе [9] авторами предложена модификация данного подхода, заключающаяся не в выявлении границ слов в конкретной фразе, а в составлении списка слов, используемых в тексте, на основе статистической обработки корпуса документов. Такой обезличенный словарь (без перевода и свойств каждого слова) может использоваться для сегментации фраз. Отличие данного подхода состоит в том, что он позволяет в первую очередь подбирать наиболее длинные слова, что позволяет сокращать неоднозначности. Например, последовательность 交换式局 может рассматриваться как два термина: 交换式 (*переключение*) и 局 (*офис, помещение, служба, ...*). В свою очередь, 交换式 распадается на слова 交换 (*обмен*) и 式 (*образец, правило, стандарт, формула?...*). Все вместе это означает *коммутиционный центр*. Если при сегментации фразы в приоритетном порядке выделять наиболее длинные последовательности, в том числе из сгенерированного обезличенного словаря, то можно существенно сократить число вариантов сегментации фразы.

Реализация предлагаемого подхода

В данной работе не преследовалась цель полного синтаксического анализа, а только апробировалась возможность извлечения фактов, поэтому был создан несложный парсер на языке SWI-Prolog объемом 200 строк, опирающийся на грамматику и небольшой словарь. Грамматика в формате Бэкуса-Наура на языке Prolog имеет вид, представленный следующим фрагментом:

```
group(sentence, [subject, predicate,
object]).
group(sentence, [subject, object,
predicate]).
group(sentence, [subject, predicate,
object, afterlog]).
group(subject, [nouns]).
group(subject, [pronouns]).
group(subject, [attribute, nouns]).
group(subject, [nouns, attribute]).
group(subject, [subject, link, subject]).
group(nouns, [noun]).
group(nouns, [noun, noun]).
...
```


Словарь содержит минимум атрибутов слов, необходимых для синтаксического разбора. В частности, для предлогов указывается, с какими частями речи они сочетаются. Приведем примеры частей речи (рис. 1). Первый аргумент предикатов частей речи содержит иероглиф(ы), второй – произношение (*пиньинь*), затем перевод на русский язык и атрибуты (число, лицо, ...).

Алгоритм синтаксического разбора предложения является рекурсивным и выглядит следующим образом.

1. Выбрать модель предложения (предикат `group(sentence,[List])` из грамматики).
2. Выбрать первый элемент из списка членов предложения `List`.
3. Отделить N (максимально возможное число) иероглифов от предложения.
4. Провести синтаксический разбор члена предложения.
5. Если разбор удачный, перейти к п. 7, иначе $N := N - 1$.
6. Если $N=0$, выбрать следующую модель предложения и перейти к п. 2.
7. Провести синтаксический анализ оставшейся части предложения.

Синтаксический разбор члена предложения выполняется рекурсивно с помощью того же самого алгоритма, но с использованием модели предложения (`sentence`), а более мелкой структурной единицы (`subject`, `object`, `predicate`, `attribute`, `noun`, `verb`, `preposition`, ...). Если на самом нижнем уровне грамматики парсер не находит словарного слова, он пытается подставить слово из обезличенного словаря, полученного путем статистической обработки корпуса документов. Естественно, слова из запроса также используются для синтаксического анализа в качестве обезличенных словарных слов. В предложенном алгоритме перебор вариантов сегментации фразы выполняется начиная с наиболее длинных последовательностей символов (п. 3 алгоритма) с целью предпочтительного выбора наиболее длинных слов.

Оценка сложности алгоритма

Пусть N – число иероглифов в предложении, b – коэффициент ветвления дерева грамматики, d – средняя глубина дерева синтаксического разбора, s – среднее число иероглифов в слове. Тогда число шагов спуска по дереву решений для парсинга первого слова составит $M_1 = \frac{(N-s)b^d}{2}$, а для каждого последующего i -го слова $M_i = \frac{(N-is)b^d}{2}$.

Таким образом, общее число шагов алгоритма M составит $M = \frac{1}{2} \sum_{i=1}^{N/s} (N-is)b^d$.

Для фразы длиной 12 иероглифов потребуется 1,44 млн шагов алгоритма. Для более длинных фраз и с более сложной грамматикой время поиска решений может стать ощутимым и, возможно, потребует применения методов редуцирования дерева решений.

Результаты экспериментов

Исследование работоспособности предлагаемого подхода проводилось на простой грамматике и небольшом словаре, сформированных на основе телевизионного курса китайского языка Д. Петрова «Китайский язык за 16 часов» (<http://16polyglot.ru/chinese/>). На рисунке 2 приведен пример синтаксического разбора фразы *我在中国饭店工作 (Я работаю в китайском ресторане)*.

Здесь каждое слово описано следующими атрибутами: иероглиф, произношение, перевод на русский язык, число, падеж (для существительных). Все слова из данного предложения присутствовали в словаре, поэтому синтаксический разбор выполнен на 100 %. Заменяем теперь в исходной фразе слово *中国 (Китай)* на *意大利 (Италия)* и получим следующую фразу: *我在意大利饭店工作*. Слово *Италия* отсутствует в словаре, поэтому оно было извлечено из списка обезличенных слов и не снабжено переводом, но атрибуты (число и падеж) взяты из свойств предлога *在 (в)* и приписаны этому обезличенному слову (см. рис. 3).

Таким образом, продемонстрирована работоспособность алгоритма синтаксического анализа китайских текстов в условиях ограниченного словаря.

Заключение

В результате проведенного исследования апробирован алгоритм синтаксического анализа китайских предложений, опирающийся на ограниченный словарь. Результаты синтаксического анализа могут использоваться для последующего извлечения фактов из текстовых документов. Поскольку предложенный алгоритм основан на переборе всех элементов грамматики и словаря, сложность алгоритма на реальных задачах может оказаться неприемлемой. Если речь идет о синтаксическом анализе для поиска и извлечения фактов из текстов, то каждую фразу в первую очередь следует проверять на наличие искоемых паттернов в составе запроса; если хотя бы один из них отсутствует, фразу можно сразу пропустить. Для редуцирования дерева поиска также можно использовать подход, предложенный в работе [10] и основанный на том, что дерево решений часто образовано повторяющимися фрагментами. Применительно к рассматриваемой задаче это может означать следующее: если разные модели предложений содержат одну и ту же часть,

```

pronoun(      '我',      'wǒ',      'я', singular, '1st').
modal_verb(   '要',      'yào',      'намереваться').
verb(         '工作',    'gōngzuò', 'работать').
adjective(    '大',      'dà',      'большой, noun').
preposition(  '在',      'zài',      'в', noun).
afterlog(     '了',      'le',      '*однократное прошедшее время', verb).
link(         '和',      'hé',      'и').

```

Рис. 1. Примеры частей речи

Fig. 1. The examples of parts of speech

```

Model: [subject, object, predicate]
  subject
    nouns
      pronoun
        我, wǒ, я, singular, 1st
  object
    preposition
      在, zài, в, singular, prepositional
    noun
      中国, zhōngguó, Китай, singular, prepositional
      饭店, fàndiàn, ресторан, singular, prepositional
  predicate
    verb
      工作, gōngzuò, работать

```

Рис. 2. Пример синтаксического разбора

Fig. 2. The syntactic analysis example

```

Model: [subject, object, predicate]
  subject
    nouns
      pronoun
        我, wǒ, я, singular, 1st
  object
    preposition
      在, zài, в, singular, prepositional
    noun
      意大利, singular, prepositional
      饭店, fàndiàn, ресторан, singular, prepositional
  predicate
    verb
      工作, gōngzuò, работать

```

Рис. 3. Пример работы алгоритма синтаксического анализа китайских текстов в условиях ограниченного словаря

Fig. 3. The example of the syntactic analysis algorithm for Chinese texts in the context of restricted dictionary

например группу подлежащего, то успешный синтаксический разбор этой группы может быть представлен в другие модели предложений, в которых группа подлежащего также присутствует.

Литература

1. Xue N. Chinese word segmentation as character tagging. Intern. Jour. Computational Linguistics and Chinese Language Processing, February 2003, vol. 8, no. 1, pp. 29–48.
2. Zeng D., Wei D., Chau M., Wang F. Domain-specific Chinese word segmentation using suffix tree and mutual information. Intern. Jour. Information Systems Frontiers. March 2011, vol. 13, iss. 1, pp. 115–125.
3. Томита-парсер. URL: <http://tech.yandex.ru/tomita/> (дата обращения: 18.11.2016).

4. Christopher M.D., Surdeanu M., Bauer J., Finkel J., Bethard S.J., and McClosky D. The Stanford CoreNLP Natural Language Processing Toolkit. Proc. 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 2014, pp. 55–60.

5. CKIP Chinese Parser. URL: <http://140.109.19.112/> (дата обращения: 21.11.2016).

6. Hsieh Yu-M., Yang D.-Ch., Chen K.-J. Improve parsing performance by self-learning. Intern. Jour. Computational Linguistics and Chinese Language Processing, June 2007, vol. 12, no. 2, pp. 195–216.

7. Yang D.Ch., Hsieh Yu-M., Chen K.-J. Resolving ambiguities of chinese conjunctive structures by divide-and-conquer approaches. URL: http://godel.iis.sinica.edu.tw/CKIP/paper/Resolving_Ambiguities_of_Chinese_Conjunctive_Structures_by_Divide_and_Conquer_Approaches.pdf (дата обращения: 21.11.2016).

8. Bessmertny I.A., Platonov A.V., Poleschuk E.A., Pen-

gyu Ma. Syntactic text analysis without a dictionary. *Proc Conf. Application of Information and Communication Technology (AICT-2016)*. 2016, pp. 100–105.

9. Бессмертный И.А., Юй Чуцяо, Ма Пенюй. Статистический метод извлечения терминов из китайских текстов без сег-

ментации фраз // Науч.-технич. вестн. информ. технологий, механики и оптики. 2016. Т. 16. № 6. С. 1096–1102.

10. Бессмертный И.А. Методы поиска информации с использованием интеллектуального агента // Изв. вузов. Приборостроение. 2009. Т. 52. № 12. С. 26–31.

Software & Systems
DOI: 10.15827/0236-235X.117.138-142

Received 23.11.16
2017, vol. 30, no. 1, pp. 138–142

AUTOMATIC SYNTACTIC ANALYSIS OF CHINESE SENTENCES BY A RESTRICTED DICTIONARY

Yu Chuqiao¹, *Postgraduate Student, yuchuqiao123@gmail.com*

I.A. Bessmertny, *Dr.Sc. (Engineering), Professor, bia@cs.ifmo.ru*

¹ *The National Research University of Information Technologies, Mechanics and Optics, Kronverksky Ave. 49, St. Petersburg, 197101, Russian Federation*

Abstract. The paper considers a problem of natural language processing of Chinese texts. One of the relevant tasks in this area is automatic fact acquisition by a query since existing automatic translators are useless for this task. The suggested approach includes a syntactic analysis of phrases and matching parts of speech founded with a formalized query.

The purpose of the study is direct fact extracting from original texts without translation. For this purpose the paper suggests to use an approach based on syntactic analysis of sentences from a text with further comparison of the found parts of speech with a formalized subject–object–predicate query. A key feature of the proposed approach is a lack of a segmentation phase of a hieroglyph sequence in a sentence by words. The bottleneck at this task is a dictionary because interpretation of a sentence is impossible without even a single word in the dictionary. To eliminate this problem the authors propose to identify a sentence model by function words while restraint of the dictionary could be compensated by automatic building of a thesaurus using statistical processing of a document corpus. The suggested approach is tested on a small topic where it demonstrates its robustness. There is also an analysis of temporal properties of the developed algorithm.

As the proposed algorithm uses a direct-search method, the parsing speed for real tasks could be unacceptably low and this is a subject for further research.

Keywords: natural language, syntactic analysis, fact extraction, thesaurus, search tree.

References

1. Xue N. Chinese Word Segmentation as Character Tagging. *Computational Linguistics and Chinese Language Processing*. 2003, vol. 8, no. 1, pp. 29–48.
2. Zeng D., Wei D., Chau M., Wang F. Domain-specific Chinese word segmentation using suffix tree and mutual information. *Information Systems Frontiers*. 2011, vol. 13, iss. 1, pp. 115–125.
3. *Tomita-parser*. 2015. Available at: <https://tech.yandex.ru/tomita/> (accessed November 18, 2016).
4. Christopher M.D., Surdeanu M., Bauer J., Finkel J., Bethard S.J., McClosky D. The Stanford CoreNLP Natural Language Processing Toolkit. *Proc. 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. 2014, pp. 55–60.
5. *CKIP Chinese Parser*. Available at: URL:<http://140.109.19.112/> (accessed November 21, 2016).
6. Hsieh Yu-M., Yang D.-Ch., Chen K.-J. Improve Parsing Performance by Self-Learning. *Computational Linguistics and Chinese Language Processing*. 2007, vol. 12, no. 2, pp.195–216.
7. Yang D.Ch., Hsieh Yu-M., Chen K.-J. Resolving Ambiguities of Chinese Conjunctive Structures by Divide-and-conquer Approaches. *IJCNLP200*. Available at: http://godel.iis.sinica.edu.tw/CKIP/paper/Resolving_Ambiguities_of_Chinese_Conjunctive_Structures_by_Divide_and_conquer_Approaches.pdf (accessed November 21, 2016).
8. Bessmertny I.A., Platonov A.V., Poleschuk E.A., Ma P. Syntactic Text Analysis Without a Dictionary. *Application of Information and Communication Technology (AICT-2016)*. 2016, pp. 100–105.
9. Bessmertny I.A., Yu Ch., Ma P. Statistical method of term extraction from chinese texts without preliminary segmentation of phrases. *Nauch.-tehnich. vestn. inform. tekhnology, mekhaniki i optiki* [Scientific and Technical Jour. of Information Technologies, Mechanics and Optics]. 2016, vol. 16, no. 6, pp. 1096–1102 (in Russ.).
10. Bessmertny I.A. Methods of information retrieval by intelligent agent. *Izv. vuzov. PriBORostroenie* [Jour. of Instrument Engineering]. 2009, vol. 52, no. 12, pp. 26–31 (in Russ.).

УДК 667.6

DOI: 10.15827/0236-235X.117.143-147

Дата подачи статьи: 19.09.16

2017. Т. 30. № 1. С. 143–147

РАЗРАБОТКА И ВНЕДРЕНИЕ МОДЕЛИ ПРОГНОЗИРОВАНИЯ ЦВЕТОВОГО ОТКЛОНЕНИЯ ПОЛИМЕРНОГО ПОКРЫТИЯ ОЦИНКОВАННОЙ ПОЛОСЫ

В.М. Осколков, аспирант, vasilii_oskolkov@outlook.com;

И.А. Варфоломеев, к.т.н., доцент, igor.varf@gmail.com;

А.Н. Виноградова, к.т.н., доцент, lnvinogradova@bk.ru;

Е.В. Ершов, д.т.н., профессор, eve@chsu.ru

*(Череповецкий государственный университет,
просп. Луначарского, 5, г. Череповец, 162600, Россия)*

Статья посвящена результатам исследования метода повышения качества оцинкованной полосы с полимерным покрытием за счет уменьшения цветового отклонения, используя методы моделирования.

Предложенная модель прогнозирования цветового отклонения состоит из трех подмоделей, каждая из которых прогнозирует одну из координат цветового пространства CIELab. Основой каждой из подмоделей является алгоритм машинного обучения Random Forest. По спрогнозированным значениям цветовых координат вычисляется полное цветовое отклонение.

Рассмотрен алгоритм построения дерева принятия решений. Описаны основные параметры, влияющие на цветовое отклонение, получаемые из трех источников: данные из сертификата о краске, характеристика входного рулона для покраски, технологические параметры.

Разработан подход быстрой и эффективной интеграции модели прогнозирования в существующую ИТ-инфраструктуру предприятия с помощью трансляции модели в БД. Разработан скрипт для трансляции модели на используемые при разработке ПО на промышленных предприятиях языки программирования SQL и .NET.

Описаны этапы трансляции модели прогнозирования с языка R на язык SQL: генерация кода, заполнение таблиц. Прогнозирование в БД осуществляется за 0,3 секунды, этого достаточно для применения в производстве в режиме реального времени.

Использование разработанной модели делает возможным прогнозирование полного цветового отклонения полимерного покрытия со средней относительной ошибкой модели 6,1 %.

Ключевые слова: интеграция моделей, генерация кода, Random Forest, регрессия, дерево решений, прогнозирование, цветовое отклонение, предварительно окрашенный прокат, полимерное покрытие, coil coating.

В настоящее время на рынках России, США и Западной Европы по-прежнему сохраняются относительно высокие цены на листовой прокат. Одним из главных направлений металлургической промышленности является производство предварительно окрашенного проката – продукции с высокой добавленной стоимостью. В связи с этим предприятия ищут новые пути для увеличения конкурентных преимуществ своей продукции. Например, «Северсталь» расширяет производство и готовит к запуску третью линию по покрытию металла, которая включает агрегат полимерных покрытий.

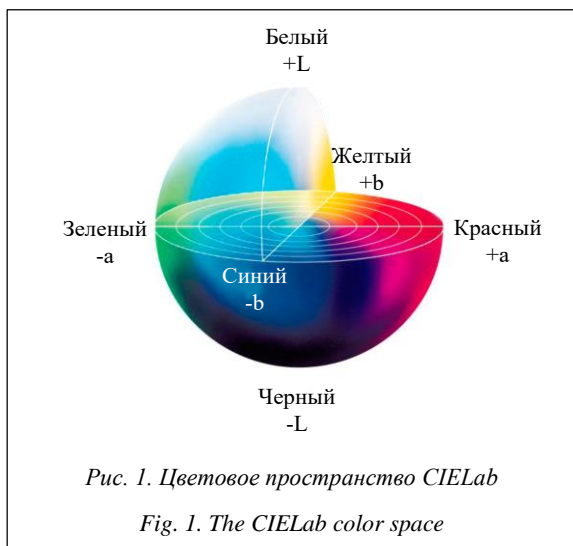
Процесс производства предварительно окрашенного проката оцинкованной полосы по технологии coil coating осуществляется на агрегате полимерных покрытий и состоит из следующих этапов: обработка поверхности полосы хромирующим раствором, нанесение грунтовочного слоя, сушка грунтовочного слоя, нанесение эмали (отделочного слоя) окрасочными валами, сушка отделочного слоя, смотка полосы в рулоны.

Компания «Тримо-ВСК» отмечает, что 5–18 % продукции проката с полимерным покрытием, поставляемой отечественными заводами, имеет отклонения от требований заказчика и норм ГОСТ, тогда как у зарубежных компаний доля такой продукции составляет в среднем 0,28 % [1]. Также от-

мечается, что партии поставляемой большинством российских компаний продукции одного цвета отличаются оттенками [1]. Поэтому так актуальна для российских металлургических предприятий задача повышения качества данной продукции.

Нанесение покрытий осуществляется с помощью грунтовочного и отделочного коутеров, которые состоят из наносящего валька, заборного валька и ванны с краской. Цвет *лакоокрасочного материала* (ЛКМ) определяется согласно каталогу цветовых гамм RAL в соответствии со стандартными образцами. Измерение цветовых координат полимерного покрытия образца выполняется с помощью спектрофотометра во время лабораторных испытаний готовой продукции.

Геометрический смысл цветового отклонения заключается в том, что оно описывает разницу между двумя цветовыми воздействиями, определяемую как евклидово расстояние между точками, описывающими эти цвета в пространстве CIELab (рис. 1) с системой координат L^* , a^* , b^* . Данное пространство разработано Международной комиссией по свету с целью упрощения обмена информацией о цвете. Оно образуется в результате представления основных цветов (белого, черного, синего, желтого, зеленого, красного) в системе координат.



Значение координаты L характеризует цветовое различие в светлую (белую) и темную (черную) стороны относительно стандарта. Положительное значение координаты a говорит о том, что цвет образца более красный, чем стандарт, отрицательное – цвет образца более зеленый. Координата b показывает отклонение по желтой и синей составляющим цвета.

Полное цветовое различие рассчитывается по формуле $\Delta E = \sqrt{\Delta L^2 + \Delta a^2 + \Delta b^2}$, где $\Delta L = L - L_0$, L – координата образца, L_0 – координата стандарта; $\Delta a = a - a_0$, a – координата образца, a_0 – координата стандарта; $\Delta b = b - b_0$, b – координата образца, b_0 – координата стандарта.

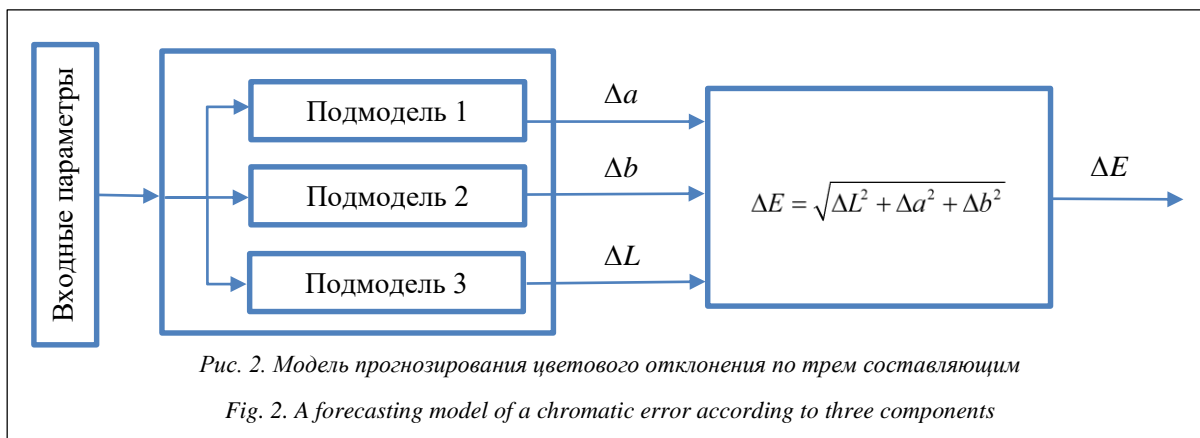
Построение модели прогнозирования

Применение моделирования сложного технологического процесса позволяет решать различные задачи по повышению качества продукции в металлургии [2]. Результаты прогнозирования можно использовать для подбора оптимальных значений технологических параметров для повышения качества продукции [3]. Для управления технологическими параметрами с целью уменьшения значения

цветового отклонения отделочного покрытия готовой продукции с помощью методов машинного обучения была построена прогнозная модель, состоящая из трех подмоделей, каждая из которых осуществляет прогнозирование одной из цветовых координат (рис. 2). По полученным координатам по вышеприведенной формуле вычисляется полное цветовое отклонение.

Входные параметры собираются из нескольких источников: данные из сертификата о краске, характеристики входного рулона для покраски, технологические данные. Краска обладает следующими свойствами: вязкость, плотность жидкой краски, плотность сухой краски, сухой остаток по объему, сухой остаток по весу, толщина сухой пленки, толщина жидкого слоя эмали, укрывистость, субстрат, приведенная толщина металла, блеск, адгезия, Т-изгиб, прочность при растяжении по Эриксену, прочность при обратном ударе, стойкость к истиранию растворителем, твердость по карандашу, начальное цветовое отклонение ЛКМ ($\Delta L'_{st}$, $\Delta a'_{st}$, $\Delta b'_{st}$, $\Delta E'_{st}$). Характеристики входного рулона для покраски: марка стали, толщина листа, температура перед входом в агрегат, толщина цинкового покрытия, теплоемкость металла.

Технологические данные собираются в процессе производства с датчиков основных узлов агрегата полимерных покрытий: разматыватель, накопитель полосы на входе, печь грунтовочного покрытия, печь отделочного покрытия, секция очистки, химический коутер, грунтовочный коутер, отделочный коутер, выходной накопитель, моталка. Технологический процесс характеризуется следующими основными параметрами: натяжение полосы, скорость полосы на каждом технологическом участке, проводимость чистящего раствора, температура чистящего раствора, температура сушильного воздуха, пиковая температура металла в печи грунтовочного покрытия, пиковая температура металла в печи отделочного покрытия, температура охлаждающей воды после печи грунтовочного слоя, температура воды после печи отделочного слоя, зазор, усилие и скорость вращения на каждом ролике, температура и скорость вра-



щения вентиляторов в каждой из семи печных зон, толщина покрытия ЛКМ. Все эти данные накапливаются по каждому метру производимой продукции.

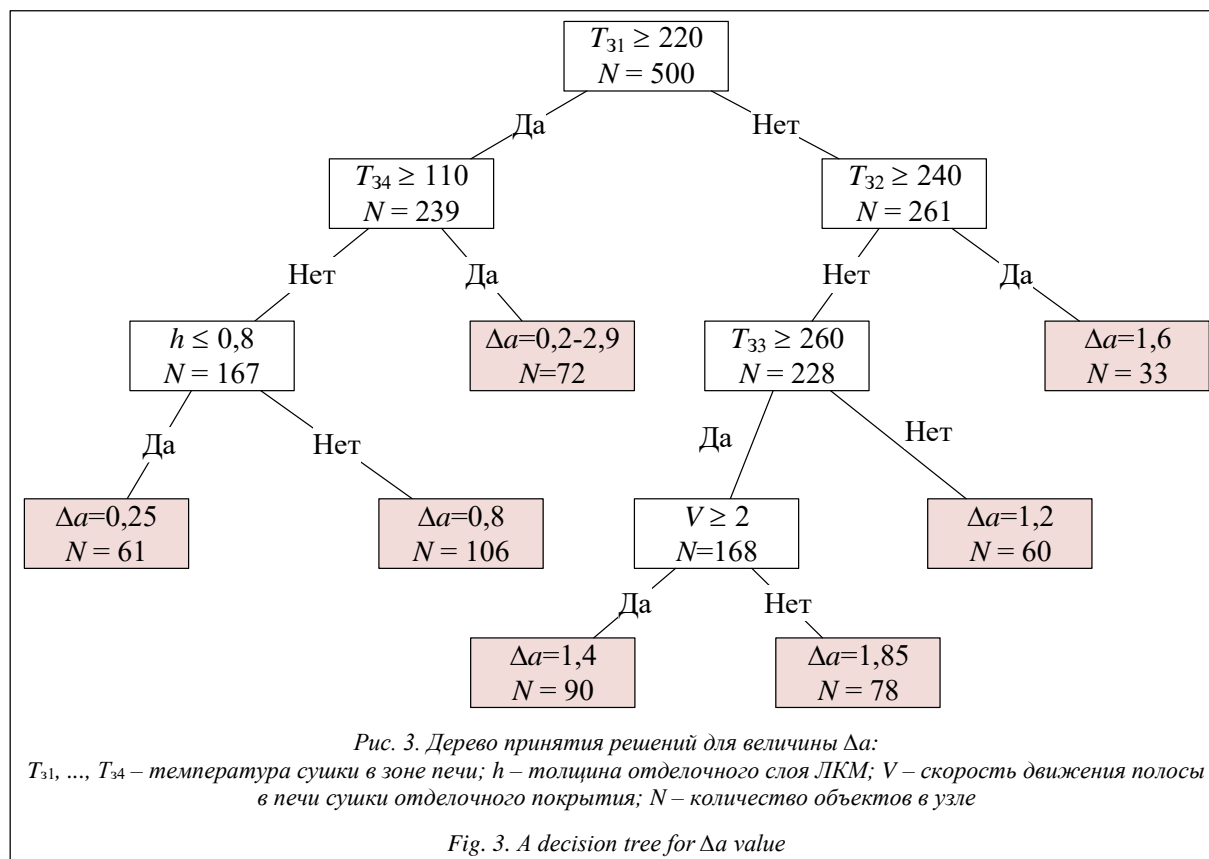
При реализации каждой из подмоделей использовался алгоритм машинного обучения Random Forest, который применяется в решении задач прогнозирования в различных отраслях [4, 5]. Данный алгоритм использует ансамбль решающих деревьев и сочетает в себе два метода – беггинга (bagging) и случайных подпространств. Дерево принятия решений представляет собой бинарное дерево. На ребрах (ветках) дерева – атрибуты, от которых зависит целевая функция, в узлах – атрибуты, по которым различаются случаи, в терминальных вершинах (листьях) – значения целевой функции. Общий принцип построения дерева решений – выбор атрибута из входных атрибутов и разбиение значений по этому атрибуту. Дальнейшее разбиение производится аналогично для каждого. Важным этапом построения является выбор критерия атрибута таким образом, чтобы в результате разбиения получаемые подмножества принадлежали к одному классу [6]. При разработке модели прогнозирования для каждой составляющей цветового отклонения автоматически строятся 300 деревьев. Расчет интервала, в котором находится результат, производится путем голосования каждого дерева из ансамбля. Пример дерева решений для прогнозирования Δa представлен на рисунке 3.

Использование описанной модели делает возможным прогнозирование полного цветового отклонения полимерного покрытия со средней относительной ошибкой модели 6,1 %. Это позволяет более гибко реагировать на требования клиента, повысить качество продукции, снизить себестоимость и сроки производства.

Внедрение модели прогнозирования цветового отклонения в систему управления технологическим процессом

Построенная модель прогнозирования решает задачу статистической обработки большого набора данных. Для реализации был выбран язык программирования R, который широко используется для статистического анализа, предсказания и визуализации данных [7]. По данным исследования, наиболее эффективной реализацией алгоритма Random Forest с точки зрения точности является его параллельная реализация parRF_t, включенная в пакет caret для языка R [8].

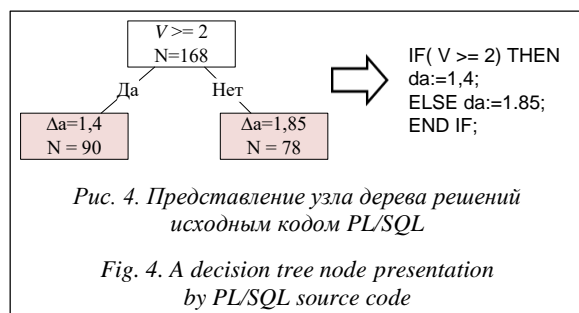
После разработки модели необходимо интегрировать ее в существующие ИТ-системы компании. Простота интеграции для крупных промышленных предприятий является важным фактором при использовании нового программного продукта. Технологические данные с агрегата записываются в БД. Исходя из этого можно сделать вывод о том, что максимально эффективным с точки зрения



ИТ-архитектуры будет внедрение модели прогнозирования на уровне БД.

Для упрощения интеграции модели был разработан скрипт на языке R, который в автоматическом режиме транслирует модель, построенную с помощью алгоритма Random Forest с языка R в программный код на языке SQL.

Каждый узел полученных деревьев может быть преобразован в конструкцию вида if-then-else. Таким образом, итоговая модель представляет собой набор соответствующих вложенных операторов языка PL/SQL (рис. 4).



Генерация SQL-кода

Генерация if-then-else конструкций приводит к большому количеству строк кода (1 дерево – 1 200 строк кода). В данном случае БД долго исполняет запрос, поэтому эффективнее транслировать деревья в записи.

На первом этапе генерируется код для создания таблицы (табл. 1) в БД со следующими полями: номер дерева, имя параметра, значение условия перехода, ссылка на узел дерева при истинном результате выполнения условия, ссылка на узел дерева при ложном результате выполнения условия, тип узла, значение для терминального узла.

На втором этапе для каждого узла всех деревьев в ансамбле генерируется конструкция INSERT для вставки записи в созданную таблицу. Обход дерева производится рекурсивно в глубину, начиная с корня дерева. Каждый узел характеризуется следующими параметрами: имя, тип узла, для нетерминальных узлов – логическое условие перехода к дочерним узлам, ссылки на дочерние узлы, для тер-

минальных узлов – результирующее значение. Поле тип узла (node_type) содержит определенное значение для каждого вида узла: 0 – корневой, 1 – внутренний, 2 – терминальный.

Таблица 1

Структура таблицы для хранения деревьев

Table 1

The structure of a storing tree table

	Имя	Тип данных	Описание
	Id	int	ID узла
	tree_num	int	Номер дерева
	param_name	nvarchar(10)	Наименование параметра
	cond_value	float	Значение условия перехода
	true_id	int	Ссылка, если ИСТИНА
	false_id	int	Ссылка, если ЛОЖЬ
	node_type	int	Тип узла
	value	int	Значение

После завершения работы программы сгенерированный код выполняется в системе управления БД Oracle для создания в ней таблицы с записями (табл. 2), соответствующими модели прогнозирования.

Прогнозирование новой записи осуществляется с помощью вызова специальной процедуры, хранимой в БД, которая вычисляет значение для каждого дерева и усредняет его. Вычисление выполняется за 0,3 секунды. Этого достаточно для использования в производстве в режиме реального времени.

Заключение

Таким образом, собрана вся информация, накапливаемая в процессе производства и доступная в сертификатах о краске. Используя эти параметры, авторы разработали модель, которая позволяет прогнозировать полное цветовое отклонение полимерного покрытия со средней относительной ошибкой модели 6,1 %. В основе модели прогнозирования лежит алгоритм Random Forest, а именно – его параллельная реализация в пакете parRF_t языка R. Кроме того, разработан подход, который

Таблица 2

Фрагмент таблицы БД, содержащей узлы дерева

Table 2

A fragment of a database table containing tree nodes

	ID	TREE_NUM	PARAM_NAME	COND_VALUE	TRUE_ID	FALSE_ID	NODE_TYPE	VALUE
1	1101	3	V178	2,8515	1102	1103	1	
2	1104	3					2	1,85
3	1100	3	V207	0,024	1101	1104	1	
4	1098	3	V212	0,00605	1099	1100	1	
5	1106	3					2	0,35
6	1107	3					2	1,4
7	1105	3	V210	0,1085	1106	1107	1	
8	1097	3	V215	0,0037	1098	1105	1	

позволяет быстро и эффективно внедрять разработанные модели прогнозирования на предприятии.

ПО «Генератор программного кода моделей Random Forest» было зарегистрировано в Роспатенте, свидетельство № 2016616932 от 22.06.2016.

Литература

1. Игнатенко Т. Оцинкованный и окрашенный прокат: нерадужные перспективы // *Металлоснабжение и сбыт*. 2015. № 4. С. 36–45.
2. Виноградова Л.Н., Шумилова Е.С. Алгоритмическое обеспечение системы прогнозирования параметров макроструктуры и качества агломерата на основе нейронной сети и фрактальной аппроксимации // *Инновационные тенденции развития российской науки: сб. матер. III Междунар. науч.-практич. конф.* 2010. С. 282–284.
3. Варфоломеев И.А., Ершов Е.В., Богачев Д.В. Оптико-

электронный контроль цветового отклонения полимерного покрытия оцинкованной полосы // *Дефектоскопия*. 2015. № 5. С. 44–51.

4. Картиев С.Б., Курейчик В.М. Алгоритм классификации, основанный на принципах случайного леса, для решения задачи прогнозирования // *Программные продукты и системы*. 2016. № 2. С. 11–15.
5. Иванова Е.И. Интеллектуальная информационная система для решения задач прогнозирования неисправностей вагонного оборудования на железнодорожном транспорте // *Программные продукты и системы*. 2015. № 4. С. 231–236.
6. Чистяков С.П. Случайные леса: обзор // *Тр. Карельского НЦ РАН*. 2013. № 2. С. 117–136.
7. Кабаков Р.И. R в действии. Анализ и визуализация данных на языке R. М.: ДМК Пресс, 2014. С. 33–36.
8. Fernández-Delgado M., Cernadas E., Barro S., Amorim D. Do we need hundreds of classifiers to solve real world classification problems? *Jour. of Machine Learning Research*. 2014, no. 15, pp. 3133–3181.

Software & Systems

DOI: 10.15827/0236-235X.117.143-147

Received 19.09.16

2017, vol. 30, no. 1, pp. 143–147

DEVELOPMENT AND IMPLEMENTATION OF A FORECASTING MODEL FOR A CHROMATIC ERROR OF GALVANIZED STRIP POLYMERIC COATING

V.M. Oskolkov¹, Postgraduate Student, vasily_osolkov@outlook.com

I.A. Varfolomeev¹, Ph.D. (Engineering), Associate Professor, igor.varf@gmail.com

L.N. Vinogradova¹, Ph.D. (Engineering), Associate Professor, lvinogradova@bk.ru

E.V. Ershov¹, Dr.Sc. (Engineering), Professor, eve@chsu.ru

¹ Cherepovets State University, Lunocharskogo Ave. 5, Cherepovets, 162600, Russian Federation

Abstract. The article presents the results of the study on quality improvement method for galvanized strip polymeric coating using modelling methods for chromatic deviation reduction. A predictive model of the chromatic deviation consisting of 3 sub-models is proposed; each sub-model predicts one CIELab color space model coordinate. Each sub-model is based on Random Forest machine learning algorithm. Full chromatic deviation output value is calculated from predictive coordinates. Each sub-model is based on Random Forest machine learning algorithm.

The paper considers a decision tree algorithm. It also describes the main parameters affecting chromatic deviation. Those parameters are received from 3 sources: paint certificate values, characteristics of an incoming strip coil for further painting, process parameters.

The authors have developed an approach for prompt and efficient integration of the mentioned forecasting model into existing IT infrastructure by model translation into a database. The developed script allows translating the model into programming languages used for industrial control systems (SQL, .NET).

The paper describes the following stages of forecasting model translation from R language into SQL language: code generation, filling the tables. Forecasting in a database takes 0,3 seconds which is enough for real time mode production. Application of the developed model allows forecasting chromatic deviation of a polymeric coating with a mean error of 6,1 %.

Keywords: model integration, code generation, random forest, regression, decision tree, prediction, chromatic aberration, pre coated rolling, polymer coating, coil coating.

References

1. Ignatenko T. Galvanized and coloured roll stock: doesn't look good. *Metallsnabzhenie i sbyt* [Metal Supply and Sales]. 2015, no. 4, pp. 36–45.
2. Vinogradova L.N., Shumilova E.S. Algorithmic support of a forecasting system for sinter macrostructure and quality parameters based on a neural network and fractal approximation. *Innovatsionnye tendentsii razvitiya rossiyskoy nauki: sb. mater. 3 Mezhdunar. nauch.-praktich. konf.* [Proc. 3rd Int. Science and Practice Conf. of Young Scientists "Innovative trends of Russian Science"]. 2010, pp. 282–284 (in Russ.).
3. Varfolomeev I.A., Ershov E.V., Bogachev D.V. Electro-optical monitoring of the chromatic error of the polymeric coating of a galvanized strip. *Defektoskopiya* [Russian Jour. of Nondestructive Testing]. 2015, no. 5, pp. 44–51 (in Russ.).
4. Kartiev S.B., Kureychik V.M. Classification algorithm based on random forest principles for a forecasting problem. *Programmnye produkty i sistemy* [Software & Systems]. 2016. no. 2, pp. 11–15 (in Russ.).
5. Ivanova E.I. Intelligent information system for rolling stock carriage equipment failures forecasting on railways. *Programmnye produkty i sistemy* [Software & Systems]. 2015. no. 4, pp. 231–236 (in Russ.).
6. Chistyakov S.P. Random forests: an overview. *Trudy KarNTs RAN* [Trans. of the Karelian Research Centre of RAS]. 2013, no. 2, pp. 117–136 (in Russ.).
7. Kabakov R.I. *R v deystvii. Analiz i vizualizatsiya dannykh na yazyke R* [R in Action. Data Analysis and Graphics with R]. Moscow, DMK Press, 2014, pp. 33–36 (in Russ.).
8. Fernández-Delgado M., Cernadas E., Barro S., Amorim D. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Jour. of Machine Learning Research*. 2014, no. 15, pp. 3133–3181.

УДК 658.52

DOI: 10.15827/0236-235X.117.148-151

Дата подачи статьи: 02.12.16

2017. Т. 30. № 1. С. 148–151

ПРОГРАММА ИДЕНТИФИКАЦИИ УСЛОВИЙ ТЕПЛООБМЕНА ДЛЯ ИЗДЕЛИЙ ПЛОСКОЙ ФОРМЫ

Б.И. Марголис, д.т.н., зав. кафедрой, borismargolis@yandex.ru
(Тверской государственный технический университет,
наб. Аф. Никитина, 22, г. Тверь, 170026, Россия)

Рассмотрена постановка задачи идентификации условий теплообмена для изделия плоской формы при несимметричном конвективно-радиационном теплообмене поверхностей изделия с окружающей средой и ограждающими поверхностями (нагревательными элементами) технологического оборудования. Сформулирована возможность решения поставленной задачи в среде программирования Matlab.

На основе стандартной функции *fmincon* в среде Matlab разработана программа, позволяющая по заданным теплофизическим характеристикам материала (коэффициентам теплопроводности, температуропроводности), параметрам конвективно-радиационного теплообмена (коэффициентам конвективной теплоотдачи и приведенным степеням черноты) и параметрам температурно-временного режима в печи отжига идентифицировать часть энергии радиационного теплообмена, попадающей с поверхности печи на изделие в каждой из зон.

Приведен пример идентификации параметров радиационного теплообмена в печи отжига листового прокатного стекла на основе программы моделирования температурного поля в среде Matlab. Рассмотрены особенности разработки программы, связанные с необходимостью учета изменяющихся начальных и граничных условий на каждом из этапов температурно-временного режима отжига изделия с помощью функций *pdebeg* и *pdebound* стандартной функции *pdepe* Matlab. Приведены программные коды функций, основной программы и результаты расчета температур поверхности ленты и степеней черноты поверхности печи.

Произведен анализ результатов работы программы, и продемонстрировано хорошее совпадение полученных степеней черноты с их физически обоснованными значениями при радиационном теплообмене в печах отжига стеклоизделий.

Показана перспективность использования стандартных функций Matlab для решения задач оптимизации режимов термообработки изделий в различных технологических процессах.

Ключевые слова: идентификация условий теплообмена, изделие плоской формы, конвективно-радиационный теплообмен, степень черноты, начальные и граничные условия, моделирование, температурное поле, температурно-временной режим, оптимизация режимов термообработки.

Определение параметров конвективно-радиационного теплообмена играет важнейшую роль в большинстве технологических процессов, связанных с термообработкой изделий [1]. Например, при описании процесса отжига листового стекла после создания математической модели температурного поля [2–6] возникает необходимость проверки совпадения результатов расчетов с экспериментальными значениями температурами характерных точек изделия (для ленты это обычно температура верхней поверхности).

Для решения этой задачи необходимо идентифицировать параметры конвективного (коэффициенты теплоотдачи для нижней и верхней поверхностей изделия α_1 , α_2 (Вт/(м²град))) и (или) радиационного (степени черноты нижней и верхней ограждающих поверхностей технологического оборудования ε_{s1} , ε_{s2}) теплообмена. Выбор определяемых параметров зависит от типа оборудования в соответствии с преобладающими механизмами внешнего теплопереноса. Различают конвективные, радиационные и конвективно-радиационные печи [7]. В данной работе приведен пример идентификации параметров радиационного теплообмена в печи отжига листового прокатного стекла на стеклозаводе «Красный май» (Тверская обл.) на основе программы моделирования температурного поля в среде Matlab, описанной в [6].

Для моделирования температурного поля используется стандартная функция *pdepe* Matlab [8, 9]. Функция типа дифференциального уравнения *pdedif* в силу исследования той же геометрической формы останется неизменной [6], тогда как функции начальных *pdebeg* и граничных условий *pdebound* изменятся следующим образом:

```
function [c,f,s]=pdedif(x,tau,t,dtdx)
global a
c = 1/a;
f = dtdx;
s = 0;

function t0=pdebeg(xv)
global tbeg x
[xm,num]=min(abs(xv-x));
t0=tbeg(num);% tbeg;

function [pl,ql,pr,qr]=pdebound(xl,tl,xr,tr,tau)
global alfa1 alfa2 es1 es2 sig Kel lambda a1 a2 a3 a4 b1
b2 b3 b4 rs1 rs2
tc1=a1+b1*tau;
tc2=a2+b2*tau;
tn1=a3+b3*tau;
tn2=a4+b4*tau;
pl = alfa1/lambda*(tc1-tl)+es1*sig/lambda*(rs1*(tn1+
+Kel)^4-(tl+Kel)^4);
ql = 1;
pr = -alfa2/lambda*(tc2-tr)-es2*sig/lambda*(rs2*(tn2+
+Kel)^4-(tr+Kel)^4);
qr = 1;
```

Это связано с необходимостью учета изменяющихся начальных и граничных условий на каждом из этапов температурно-временного режима отжига изделия при его перемещении в следующую зону печи отжига. В приведенных выше функциях a_1, a_2, a_3, a_4 (°C) и b_1, b_2, b_3, b_4 (°C/мин.) – начальные температуры и скорости изменения температур среды и печи соответственно под и над лентой для каждой из зон отжига.

Вспомогательной при идентификации условий теплообмена является приведенная далее функция *Iден*, определяющая отклонение рассчитанной по модели температуры верхней поверхности ленты от экспериментально полученной tl_iden :

```
function y=Iден(par,m,tau)
global l a1 b1 a2 b2 a es1 es2 sig Kel lambd alfa1 alfa2
a3 b3 a4 b4 tbeg x tl_iden t rs1 rs2
alfa1=par(1); alfa2=par(2); rs1=par(3); rs2=par(4);
options=odeset('RelTol',1e-4);
sol = pdepe(m,@pdedif,@pdebeg,@pdebound,x,tau,
options);
t = sol(:,1);
tfin=t(end,:);
y=abs(tfin(end)-tl_iden);
```

Исходные данные для решения задачи идентификации, описывающие условия в цехе отжига листового прокатного стекла, приведены в таблице 1.

Таблица 1

Исходные данные по зонам печи отжига

Table 1

Initial data by annealing furnace zones

Зона от- жига	Коор- дината печи, м	Темпера- тура среды $a_1=a_2$, °C	Темпера- тура печи $a_3=a_4$, °C	Температура поверхности tl_iden , °C
1	1,8	479	540	477
2	3,6	510	560	496
3	5,7	518	560	513
4	7,5	537	550	531
5	10,5	539	550	532
6	13,5	541	550	533
7	15,9	532	540	526
8	18,2	507	530	500
9	20,7	480	530	473
10	24,3	453	520	446
11	26,2	439	439	430
12	29,9	423	423	413
13	31,6	388	388	371
14	36,4	366	366	352
15	38,8	346	346	329
16	43,3	325	325	306
17	48,4	301	301	284
18	50,8	285	285	260

Скорость движения изделия в печи составляет $V=1,2$ (м/мин.), поэтому по данным таблицы можно рассчитать время нахождения в зоне отжига и скорости изменения температур среды и печи на каждом этапе температурно-временного режима. Для упрощения расчетов рассмотрена симметричная задача конвективно-радиационного теплообмена,

то есть $\alpha_1=\alpha_2$ и $\varepsilon_{s1}=\varepsilon_{s2}$. Данные таблицы 1 в среде Matlab удобно читать из файла Excel. При расчете использованы следующие значения параметров: $a=0,25386$ см²/мин.; $\lambda=0,008856$ Вт/см×град.; $N=6$; $\ell=0,6$ см; $\varepsilon_{s1}=\varepsilon_{s2}=0,85$; $\alpha_1=\alpha_2=0,001506$ Вт/см²×град.; $t_0(r)=444,0; 457,7; 468,0; 474,0; 475,1; 471,1; 461,7$ °C.

Для идентификации параметров теплообмена в среде Matlab необходимо использовать функцию *fmincon*, позволяющую эффективно решать задачи нелинейной оптимизации функции нескольких переменных с ограничениями типа неравенств и равенств. Приведем текст основной программы с использованием функции *fmincon*:

```
clear
close all
clc
global l a1 b1 a2 b2 a es1 es2 sig Kel lambd alfa1 alfa2
a3 b3 a4 b4 tbeg x tl_iden t rs1 rs2
TTR = xlsread('TBP непуод отжига');
Nzones=size(TTR,2)
yst2=TTR(3,:); yst1=yst2; yst4=TTR(7,:);
yst3=yst4;% среда и нагреватели слева, справа
tl_exp=TTR(1,:);
coord=TTR(9,:); L=coord(Nzones); % координаты и
длина (м)
tau_finish=TTR(10,Nzones); V=L/tau_finish % время
(мин); скорость ленты (м/мин)
tau_end=coord/V; % текущее время конца этапа
TBP (мин)
tau_zone(1)=tau_end(1); Vyst1(1)=0; Vyst2(1)=0;
Vyst3(1)=0; Vyst4(1)=0;
for i=2:Nzones
    tau_zone(i)=tau_end(i)-tau_end(i-1); % длительно-
    сти этапов TBP (мин)
    Vyst1(i)=(yst1(i)-yst1(i-1))/tau_zone(i);
    Vyst2(i)=(yst2(i)-yst2(i-1))/tau_zone(i);
    Vyst3(i)=(yst3(i)-yst3(i-1))/tau_zone(i);
    Vyst4(i)=(yst4(i)-yst4(i-1))/tau_zone(i);
    % скорости изменения температур среды и
    нагревателей на этапах (град/мин)
end;
m = 0; sig=5.67e-12; Kel=273.15; es2=0.85; es1=es2;
rs1=0.8; rs2=rs1;
alfa2=0.001506; alfa1=alfa2; a=0.6*0.4231;
lambd=0.008856;
l=0.6; N=7; x = linspace(0,l,N); Ntau=20;
N1=round(N/2);
tbeg=TTR(14,1:N);
par_prev=[alfa1; alfa2; rs1; rs2]; TolX=0.001;
A=[]; b=[]; nonlcon=[]; optset=[];
lb=[0.0015; 0.0015; 0.05; 0.05]; ub=[0.001506;
0.001506; 0.95; 0.95];
Aeq=[0 0 1 -1]; beq=[0]; par_all=[];
for i=1:Nzones
    tau_fin=tau_zone(i);
    tau = linspace(0,tau_fin,Ntau);
    if i==1
        a1=yst1(i); a2=yst2(i); a3=yst3(i); a4=yst4(i);
    else
        a1=yst1(i-1); a2=yst2(i-1); a3=yst3(i-1);
    a4=yst4(i-1);
    end;
    b1=Vyst1(i); b2=Vyst2(i); b3=Vyst3(i); b4=Vyst4(i);
```

```

par0=par_prev; t1_iden=t1_exp(i);
optset=optimset('Display','iter','TolX','TolX')
[par,fval]=fmincon(@Iden,par0,A,b,Aeq,beq,lb,ub,nonlcon,optset,m,tau);
par_all=[par_all; par'];
% options=odeset('RelTol',1e-4);
% sol = pdepe(m,@pdedif,@pdebeg,@pdebound,
x,tau,options);
% t = sol(:, :, 1)
t1=[t(:,1) t(:,N1) t(:,N)]
if i==1
    surf(x,tau,t)
    title('Температурное поле в период отжига')
    xlabel('Координата x, см')
    ylabel('Время \tau, мин')
    grid on
    hold on
else
    figure(1)
    surf(x,tau+tau_end(i-1),t)
end;
tbeg=t(end,:);
par_prev=par;
figure(2)
plot(x,t(end,:), 'r')
title(strcat('Решение при \tau = ',
num2str(tau_end(i)), '(мин)'))
xlabel('Координата x, см')
ylabel('t(x, \tau)')
grid on
hold on
end;

```

Рассчитанное температурное распределение приведено на рисунке, а идентифицированные значения параметров теплообмена и температур поверхности пластины – в таблице 2.

Из таблицы 2 видно, что отклонения рассчитанных температур верхней поверхности пластины от ее экспериментальных значений составляют десятые доли градуса, за исключением первой зоны,

что можно объяснить неравномерностью условий теплообмена на входе в печь отжига. Полученные степени черноты ϵ_{s2} представляют часть энергии радиационного теплообмена, попадающей с поверхности печи на изделие в каждой из зон. Их значения в диапазоне от 0,5 до 0,95 хорошо согласуются с физическими представлениями о радиационном теплообмене в печах отжига стеклоизделий.

Таблица 2

Результаты идентификации условий теплообмена

Table 2

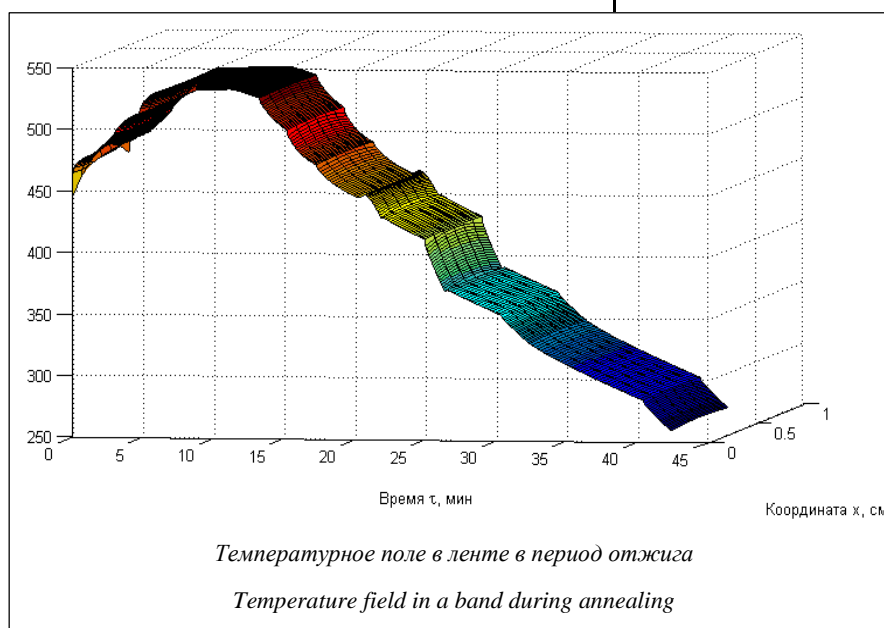
Heat-exchange condition identification results

Зона отжига	Экспериментальная температура $t1_iden$, °C	Рассчитанная температура $tfin(end)$, °C	Степень черноты поверхности печи ϵ_{s2}
1	477	479,2	0,7394
2	496	496,4	0,7722
3	513	512,8	0,8083
4	531	530,9	0,9207
5	532	531,8	0,9106
6	533	533,0	0,9153
7	526	526,0	0,9075
8	500	500,2	0,8101
9	473	472,5	0,7080
10	446	446,0	0,6423
11	430	430,0	0,7528
12	413	413,0	0,8820
13	371	371,0	0,5973
14	352	351,9	0,8225
15	329	329,0	0,7119
16	306	306,0	0,7414
17	284	284,1	0,7389
18	260	259,8	0,5361

Представленные результаты идентификации условий конвективно-радиационного теплообмена

с окружающей средой и ограждающими поверхностями можно использовать для оптимизации режимов термообработки изделий плоской формы с целью сокращения длительности технологического процесса или энергетических затрат на него [10].

Наличие в Matlab наряду с *pdepe* функции *fmincon*, решающей задачи оптимизации функции нескольких переменных с ограничениями, позволяет эффективно использовать ее для оптимизации режима отжига стеклоизделий плоской формы.



Литература

1. Рубанов В.Г. Автоматизация и управление объектами промышленности строительных материалов // Строительные материалы. 1996. № 2. С. 180–19.
2. Михеев М.А., Михеева М.И. Основы теплопередачи. М.: Энергия, 1977. 344 с.
3. Мазурин О.В., Лалыкин Н.В. Математическая модель процесса отжига листового стекла // Стекло и керамика. 1984. № 1. С. 13–15.
4. Лыков А.В. Теория теплопроводности: учеб. пособие. М.: Высш. школа, 1967. 599 с.
5. Gardon R. Calculation of temperature distributions in glass

plates undergoing heat-treatment. J. Amer. Ceram. Soc., 1958, vol. 41, no. 6, pp. 200–209.

6. Марголис Б.И. Программы моделирования температурных полей в изделиях плоской формы // Программные продукты и системы. 2016. № 2. С. 124–127.
7. Мазурин О.В., Белоусов Ю.Л. Отжиг и закалка стекла: учеб. пособие. М.: Изд-во МИСИ и БТИСМ, 1984. 114 с.
8. Лазарев Ю. Моделирование процессов и систем в MATLAB: учеб. курс. СПб: Питер, 2005. 512 с.
9. Дьяконов В.П. MATLAB 7.*/ R2006/ R2007: самоучитель. М.: ДМК Пресс, 2008. 768 с.
10. Марголис Б.И. Нахождение оптимального режима отжига стеклоизделий, обеспечивающего минимальные энергозатраты // Стекло и керамика. 2003. № 5. С. 12–13.

Software & Systems

DOI: 10.15827/0236-235X.117.148-151

Received 02.12.16

2017, vol. 30, no. 1, pp. 148–151

PROGRAM FOR HEAT CONDITIONS IDENTIFICATION IN FLAT PRODUCTS

B.I. Margolis¹, Dr.Sc. (Engineering), Head of Chair, borismargolis@yandex.ru

¹ Tver State Technical University, Nikitin Quay 22, Tver, 170026, Russian Federation

Abstract. The article considers the problem statement on identification of heat transfer conditions for flat products with asymmetrical shapes convective-radiative heat transfer surfaces to the environment and enclosing surfaces (heating elements) of the process equipment. The paper formulates a possible solution of the problem in Matlab.

The program, which has been developed based on the standard *fmincon* function in MatLab, allows identifying the part of radiation heat transfer energy, which gets from a furnace surface to the product in each zone. For this purpose it uses predetermined thermo-physical characteristics of the material (thermal conductivity, thermal diffusivity), the parameters of convective-radiative heat transfer (coefficients of convective heat transfer and an emissivity factor) and the temperature and time parameters on an annealing furnace.

The paper presents an example of radiative heat transfer parameter identification in an annealing furnace for rolled glass sheet on the basis of the temperature field simulation program in Matlab. The authors consider the features of program development related to the need to take into account changing initial and boundary conditions at each stage of the temperature-time mode of an annealing product using *pdebeg* and *pdebound* of the standard functions *pdepe* in MatLab. There are software codes of functions and main program, as well as the results of calculating band surface temperatures and emissivity factors of a furnace surface.

There is the analysis of the results of the program. The paper demonstrates good agreement between the obtained emissivity factors and their physically reasonable values for radiative heat transfer in glass annealing furnaces. The paper shows the prospects of using standard Matlab functions to solve product heat treatment mode optimization problems in various technological processes.

Keywords: heat transfer conditions identification, flat shape product, convective-radiant heat transfer, emissivity factor, initial and boundary conditions, simulation, temperature field, temperature-time mode, heat treatment optimization.

References

1. Rubanov V.G. Automation and management of industry facilities of construction materials. *Stroitelnye materialy* [Construction Materials]. 1996, no. 2, pp. 18–19 (in Russ.).
2. Mikheev M.A., Mikheeva M.I. *Osnovy teploperedachi* [Heat-Transfer Principles]. Moscow, Energiya Publ., 1977, 344 p.
3. Mazurin O.V., Lalykin N.V. A mathematical model of a flat glass annealing process. *Steklo i keramika* [Glass and Ceramics]. Moscow, Stroyizdat Publ., 1984, vol. 41, iss. 1, pp. 9–13 (in Russ.).
4. Lykov A.V. *Teoriya teploprovodnosti* [Thermal Conductivity Theory]. Study guide, Moscow, Vysshaya shkola, 1967, 599 p.
5. Gardon R. Calculation of temperature distributions in glass plates undergoing heat-treatment. *Jour. Amer. Ceram. Soc.* 1958, vol. 41, no. 6, pp. 200–209.
6. Margolis B.I. Simulator programs for temperature fields in flat form products. *Programmnye produkty i sistemy* [Software & Systems]. 2016, no. 2, pp. 124–127 (in Russ.).
7. Mazurin O.V., Belousov Yu.L. *Otzhig i zakalka stekla* [Glass Flashing and Toughening]. Study guide, Moscow, MISI i BTISM Publ., 1984, 114 p.
8. Lazarev Yu. *Modelirovanie protsessov i sistem v MATLAB* [Modeling processes and systems in MATLAB]. Training course, St. Petersburg, Piter Publ., 2005, 512 p.
9. Dyakonov V.P. *MATLAB 7.*/ R2006/ R2007: samouchitel* [MATLAB 7.*/ R2006/ R2007: Teach Yourself]. Moscow, DМК Press, 2008, 768 p.
10. Margolis B.I. An Optimum Regime for Glass Annealing with Minimum Energy Consumption. *Steklo i keramika* [Glass and Ceramics]. Moscow, Stroyizdat Publ., 2003, vol. 60, iss. 5, pp. 12–13 (in Russ.).

УДК 371.693.4

DOI: 10.15827/0236-235X.117.152-156

Дата подачи статьи: 08.11.16

2017. Т. 30. № 1. С. 152–156

**ПРОФЕССИОНАЛЬНАЯ ПОДГОТОВКА ЛИЧНОГО СОСТАВА КОРАБЛЯ
В СИСТЕМЕ ЖИЗНЕННОГО ЦИКЛА ПРОДУКЦИИ ВОЕННОГО НАЗНАЧЕНИЯ***Е.Ф. Лосев, д.в.н., профессор, капитан 1-го ранга в отставке, losev1947@mail.ru**(Филиал ВУНЦ ВМФ «Военно-морская академия» в г. Калининграде,
Советский просп., 82, г. Калининград, 236036, Россия);**И.В. Кузнецов, к.в.н., капитан 2-го ранга, 89817190035@mail.ru**(Военный институт дополнительного профессионального образования
ВУНЦ ВМФ «Военно-морская академия»,**Малоохтинский просп., 80/2, г. Санкт-Петербург, 195112, Россия);**А.А. Бавула, капитан 2-го ранга, адъютант, bavel_@mail.ru;**И.А. Бурик, капитан 1-го ранга, начальник кафедры, bavel_@mail.ru**(Филиал ВУНЦ ВМФ «Военно-морская академия» в г. Калининграде,
Советский просп., 82, г. Калининград, 236036, Россия)*

В статье рассматривается профессиональная подготовка личного состава в системе управления жизненным циклом продукции военного назначения на основе так называемых сквозных контрактов жизненного цикла.

Авторами предложена концептуальная модель управления жизненным циклом продукции военного назначения с включением в единое информационное пространство жизненного цикла продукции военного назначения – профессиональной подготовки личного состава боевой части связи надводных кораблей ВМФ.

Приводится критический аспект комплексного тренажера корабельных связистов «Племя-С».

Предлагается новый подход с использованием тренажерной подготовки персонала с применением имитационной, виртуальной среды обучения для специалистов ВМФ при переходе на контракты сквозного жизненного цикла.

Авторы предполагают, что виртуальная среда моделирования позволит обучаемым приобретать уникальные навыки действий в самых разнообразных нештатных ситуациях, создать которые традиционным способом в процессе обучения не всегда представляется возможным.

Ключевые слова: профессиональная подготовка, жизненный цикл продукции военного назначения, моделирование виртуальной среды обучения.

Сложная международная обстановка, участвовавшие вооруженные конфликты и международный терроризм, развязанная информационная война против России ставят перед Министерством обороны новые задачи по совершенствованию боеготовности армии и флота. Прежде всего это касается управления и взаимодействия подразделений вооруженных сил, оснащения их современными средствами вооружения и связи.

Задачи обеспечения заданной боевой готовности ВМФ можно успешно решить только при эффективном функционировании системы военно-морского образования и системы боевой подготовки. В свою очередь, эффективность образования и боевой подготовки зависит от наличия учебно-тренировочных средств (УТС), соответствующих современному состоянию ВМФ [1].

В своем развитии УТС прошли несколько этапов. Наиболее значимый связан с внедрением современных информационных технологий [2]. Образуясь, современные информационные технологии открывают окно в потаенный мир человеческой фантазии, обеспечивая возможность моделирования «изображаемой художником» реальности в ее временном развитии [3].

В современных условиях морского боя роль вооружения, военной и специальной техники (ВВСТ) связи постоянно растет – с каждым разом на новом технологическом уровне. Связь остается ключевой

основой управления. Совершенствование технической готовности средств связи, а также профессионального обучения личного состава современным наукоемким комплексам связи, установленным на кораблях ВМФ, – одна из основных задач боевой подготовки кораблей и соединений ВМФ РФ.

В 2013 году Министерство обороны Российской Федерации (МО РФ) предложило новую систему по поддержанию технической готовности вооружений и военной техники, по ее ремонту и техническому обслуживанию. Предполагается заключение специального контракта, так называемого контракта жизненного цикла продукции военного назначения, с оборонным предприятием. Суть этого подхода в том, что производитель по заказу МО РФ не только проектирует, производит, вводит в эксплуатацию, но и поддерживает техническое состояние ВВСТ в течение всего жизненного цикла вплоть до момента списания и утилизации.

В соответствии с рекомендациями МО РФ отечественная система управления жизненным циклом для различных образцов ВВСТ, безусловно, будет иметь свою специфику, однако базовые нормативные документы, специальное ПО, методики сбора, обработки и представления информации должны быть построены на единых принципах и стандартах.

Анализ зарубежного опыта применения данных контрактов свидетельствует об их несомненной эф-

фektivности. Так, например, в военно-морских силах Великобритании за невыполнение контракта технического обслуживания в течение жизненного цикла предусматривается не только отказ от уплаты, но и наложение высоких штрафных санкций за снижение технической готовности вооружения и боевой готовности корабля в целом. Это заставляет оборонные предприятия скрупулезно выполнять взятые на себя обязательства по контракту, разрабатывать и производить высоконадежное корабельное вооружение [4].

Прогноз внедрения и эволюционное развитие контрактов жизненного цикла позволят повысить эффективность технического обеспечения ВВСТ, снизить нагрузку на бюджет МО РФ. Есть и другие очевидные преимущества. МО освобождается от второстепенных задач, получает возможность сосредоточиться в большей степени на боевой подготовке, то есть на выполнении своих главных функций.

Схематично все элементы управления жизненным циклом продукции военного назначения можно представить в виде концептуальной модели (рис. 1).

В рамках контрактов жизненного цикла продукции появляется возможность организовать обратную связь с участниками данного процесса с помощью онлайн среды – единого информационного пространства жизненного цикла продукции военного назначения.

Это позволит МО и предприятиям ОПК владеть реальной технической информацией на всех этапах

жизненного цикла поставляемой продукции и быть уверенными, что она правильно применяется и эксплуатируется. Всю получаемую онлайн информацию после критического анализа необходимо аккумулировать и использовать для усовершенствования, модернизации образцов средств связи, а также для разработки интерактивных технических инструкций и руководств по применению и использованию на любых разрешенных и учтенных мобильных носителях информации.

Изучив все элементы модели управления жизненным циклом продукции военного назначения, авторы считают целесообразным обратить внимание на то, в какой мере будет учтена возможность профессиональной подготовки (переподготовки) личного состава к обслуживанию новых систем вооружения в рамках будущих контрактов.

В настоящее время вводятся в эксплуатацию новые корабельные комплексы связи. Следовательно, возникает необходимость в квалифицированной подготовке специалистов по их применению и обслуживанию. При этом существенно повышается роль технических средств обучения личного состава. Заказчикам сложной, наукоемкой и дорогостоящей техники связи необходимо требовать от производителей одновременной поставки тренажерных комплексов, обеспечивающих освоение ВВСТ и профессиональное обучение персонала на тренажерной базе, а не на боевой технике с целью исключения поломок аппаратуры связи при неумелой эксплуатации на первоначальном этапе ее освоения.

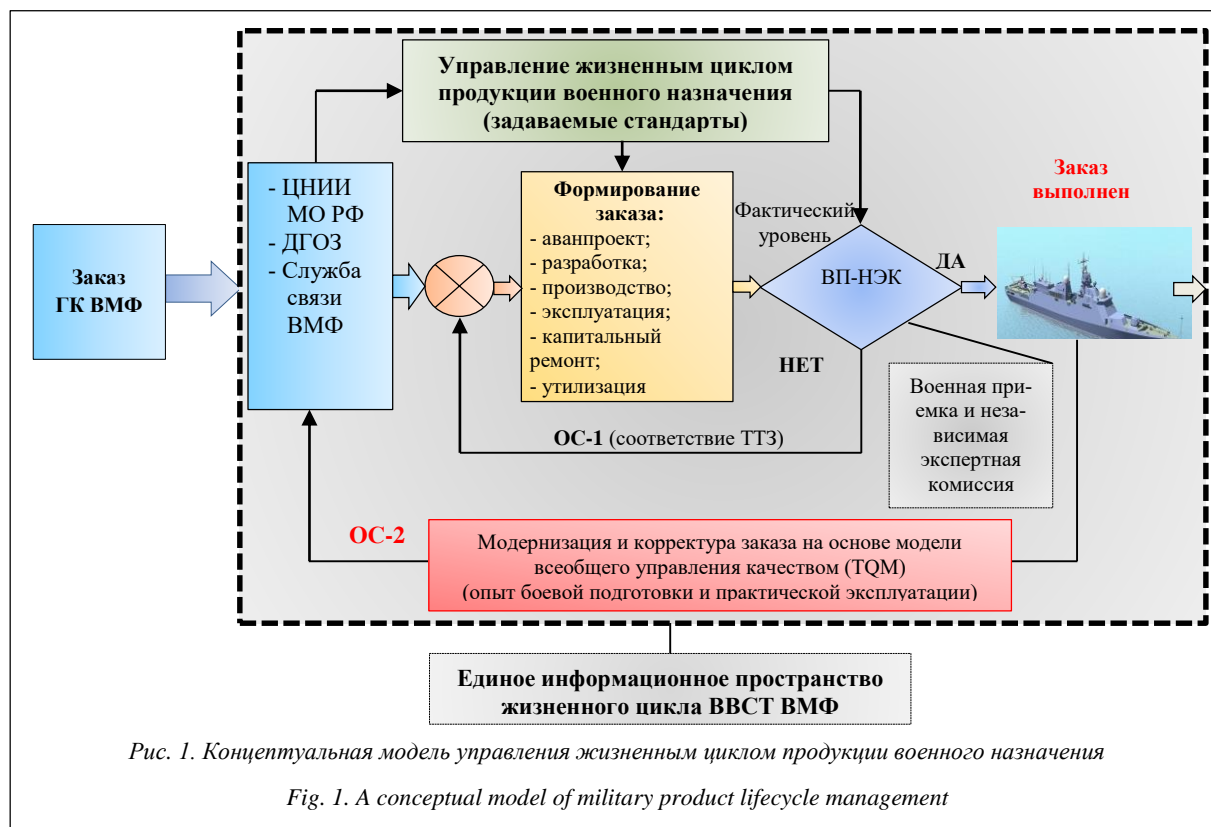


Рис. 1. Концептуальная модель управления жизненным циклом продукции военного назначения

Fig. 1. A conceptual model of military product lifecycle management

В 2005 году был разработан комплексный тренажер корабельных связистов (КТ КС) «Племя-С» (рис. 2). Данный тренажер позволил осуществлять комплексную подготовку специалистов связи надводных кораблей ВМФ по использованию автоматизированных комплексов связи. Априори тренажер, принятый на вооружение более десяти лет назад, сегодня имеет ряд недостатков, а также требует доработки и модернизации. О данных процессах пишут многие авторы (например [5]).



Рис. 2. КТ КС «Племя-С»

Fig. 2. An integrated simulator of ship's communicators "Plemya-S"

Выявленные недостатки объясняются тем, что компьютерные тренажеры являются сложными открытыми развивающимися системами с активными элементами. Особенность разработки таких систем в том, что их легче изготовить и ввести в действие, начиная с некоторого уровня сложности, а далее преобразовывать и изменять, чем отобразить формальной моделью на этапе проектирования [5, 6].

По мнению авторов, так называемая оболочка тренажера должна быть единой для всех комплексов связи, а меняться будет только ПО тренажера в соответствии с новыми образцами техники.

Известно, что разработчики компьютерных тренажеров часто возлагают на них множество функций отображения обучающей информации, не уделяя при этом должного внимания необходимой степени автоматизации функций управления обучением [7].

Предлагаемые сейчас сценарии компьютерных обучающих программ (КОП) не соответствуют сценариям учебных занятий, принятым в учебных заведениях ВМФ.

Сценарии современных КОП – это просто рядок предъявления кадров с учебной информацией. У них нет четкой структуры, в них не определены действия участников процесса обучения. Сценарии содержат кадры с темой и учебными вопросами, кадры с изучаемым материалом и с контрольными вопросами, но в них нет кадров с резюме после каждого изученного учебного вопроса, нет реаль-

ной имитации (динамики) в работе на современных комплексах связи с обратной реакцией – алгоритмом последовательности и правильности действий обучаемого. Отсутствуют заключительные кадры с выводами по представленному материалу [8, 9].

Сейчас мы имеем уникальную возможность включить обеспечение профессиональной подготовки личного состава в процесс заключения контрактов жизненного цикла ВВСТ. Актуальность проблемы продиктована изменениями, происходящими на флоте: оснащением объединений, соединений, воинских частей, кораблей ВМФ новой аппаратурой, современными телекоммуникационными комплексами связи, а также проблемными вопросами, возникающими при их применении в подразделениях ВМФ, например, выход из строя техники связи, устранение неисправностей, возникновение нештатных ситуаций, получение неудовлетворительных оценок при отработке задач по связи и многое другое.

В настоящее время практически отсутствует обратная связь, необходимая для анализа качества профессиональной подготовки корабельных связистов, не происходит адекватной корректировки (так называемой подстройки) учебного материала в соответствии с новыми образцами ВВСТ, а также отсутствуют современные средства обучения и тренажа. Чаще всего обучение осуществляется на боевой технике. При этом расходуется ее боевой ресурс, зачастую приводящий к непреднамеренному выходу из строя, что в целом снижает техническую готовность средств связи и, как следствие, боеготовность корабля. Возникла необходимость создания специализированных средств обучения и тренажа на основе элементов имитационного моделирования и виртуальной среды обучения, позволяющих с помощью компьютеров достаточно просто реализовать ситуационный подход в подготовке личного состава. Привлекает возможность создания самых разнообразных тренажерных сценариев. Такие средства способны сформировать индивидуальную траекторию подготовки каждого связиста, а впоследствии и восстановить утраченные им навыки. В интерактивной виртуальной среде обучения само обучение проходит более эффективно и интересно.

Виртуальная реальность – это новая технология неконтактного информационного взаимодействия, реализующая с помощью комплексных мультимедиа-операционных сред иллюзию непосредственного вхождения и присутствия в реальном времени в стереоскопически представленном экранном мире. Технология неконтактного информационного взаимодействия, реализуемая системой «Виртуальная реальность», позволяет компьютеру отобразить непосредственно в цифровой форме импульсы от информационной перчатки (интерфейс-перчатка) и информационного костюма. Рука пользователя, одетая в информационную перчатку, может

быть спроецирована в виртуальной форме в трехмерной компьютерно-генерированной среде. Манипулируя информационной перчаткой, пользователь может взаимодействовать с виртуальным миром, передвигая объекты, управляя ими, может также использовать набор жестов в качестве команд. При наличии информационного костюма, информационной перчатки и информационных очков со встроенными стереоскопическими экранами (очки-телемониторы) пользователь может, образно выражаясь, шагнуть прямо в виртуальный мир [3].

Тенденция подготовки персонала с использованием тренажерных комплексов является общемировой. В ряде зарубежных стран до 80 % времени всей подготовки военнослужащих отводится для занятий на специальных тренажерных комплексах. Необходимо отметить, что в ВМФ и МО РФ все больше внимания уделяется подготовке военнослужащих с применением УТС обучения, постоянно увеличивается парк тренажеров, тренажерных комплексов и КОП.

С появлением технологической возможности создания виртуальной среды обучения УТС обучения приобрели большую возможность имитационного моделирования и гибкость при смене учебных сценариев. Сегодня, в век глобальной компьютеризации, нет необходимости убеждать, что создание виртуальной среды обучения обусловлено эффективностью учебного процесса: при формировании знаний – теоретическая подготовка, а также на этапах формирования умений и навыков – тренажерная подготовка специалистов ВМФ.

Такая форма обучения позволит снизить эксплуатационную нагрузку на боевые образцы средств связи, значительно экономя их боевой ресурс (рис. 3).

Необходимо отметить, что виртуальная среда обучения не предполагает замену практической работы личного состава, а только дополняет и расширяет возможности обучения и поддерживает необходимый уровень приобретенных знаний, умений

и навыков. По сравнению с обучением специалистов на боевой технике основным преимуществом такой подготовки является возможность отработки навыков, которые невозможно получить при обучении на действующей аппаратуре. Речь идет о нештатных ситуациях: отказ техники, возникновение аварий, пожаров и тому подобное, которые можно моделировать только на УТС.

Таким образом, специалисты, прошедшие обучение с применением УТС, имеют возможность приобрести уникальные навыки действий в самых разных нештатных и экстремальных ситуациях.

Включение процесса профессиональной подготовки личного состава в контракты жизненного цикла изделий должно носить профессиональный характер. Должны разрабатываться УТС, способные имитировать реальную среду применения средств связи в максимальной степени. Разрабатывая интерактивные учебные материалы и технические руководства, необходимо привлекать специалистов в области эргономики и электронного учебного дизайна, а также педагогов, способных более эффективно структурировать учебный материал, сопровождающий познавательный процесс.

Предполагается, что такой подход позволит

- существенно снизить стоимость обучения по сравнению с традиционной подготовкой специалистов на боевой технике;
- сократить боевой расход ресурса дорогостоящей техники, а также расходных материалов;
- сократить срок подготовки специалистов, что особенно актуально в связи с сокращением срока службы по призыву до одного года;
- повысить уровень боевой подготовки за счет внедрения контрактов, основанных на управлении жизненным циклом продукции.

Кроме этого, контракт на управление жизненным циклом продукции позволяет выстроить другой уровень взаимоотношений между производителем и заказчиком. Как правило, это более долгосрочные и прочные взаимоотношения партнерства.



Они позволяют в некоторой мере защитить интересы производителя от покушений конкурентов на его заказчика.

Считаем, что контракты, основанные на управлении жизненным циклом продукции военного назначения, повысят боевую готовность корабля в целом, улучшат качество профессиональной подготовки обслуживающего персонала и эксплуатации поставляемых комплексов связи в ВМФ, существенно сузят круг посредников. Переход на данные контракты позволит более эффективно использовать огромные средства, выделенные на обеспечение обороноспособности нашей страны.

Литература

1. Захаров В.Л., Ильин В.А., Кушнарев А.Г. Оперативно-тактическая система ВУНЦ ВМФ. Какой ей быть? // Оборонный заказ. 2010. № 28.
2. Довженко В.Н., Стручков А.М., Туровский О.М. Ис-

пользование современных информационных технологий в системе подготовки кадров для ВМФ // Морской сборник. 2009. № 12. С. 44–51.

3. Роберт И.В. Современные информационные технологии в образовании. М.: Изд-во ИИО РАО, 2010. 107 с.
4. Бакарджиева С. Контракты ЖЦИ как стимул к модернизации // Умное производство. 2013. № 3. С. 65–67.
5. Сергеев В.В., Пучко Е.В., Родионов А.В. Облик перспективных тренажеров корабельных связистов // Программные продукты и системы. 2015. № 1. С. 92–99.
6. Волкова В.Н., Денисов А.А. Теория систем. М.: Высш. шк., 2006. 511 с.
7. Печников А.Н., Ветров Ю.А. Проектирование и применение компьютерных технологий обучения. Ч. 1. Концепция и применение компьютерных технологий обучения. СПб: Изд-во БГТУ, 2003. Кн. 1. 195 с.
8. Методическое руководство по разработке сценарных материалов для автоматизированных учебных занятий с использованием комплекса программ инструментальных средств «MediaTor». СПб, 2002.
9. Башмаков А.И., Башмаков И.А. Разработка компьютерных учебников и обучающих систем. М.: Филинь, 2003. 616 с.

Software & Systems

DOI: 10.15827/0236-235X.117.152-156

Received 08.11.16

2017, vol. 30, no. 1, pp. 152–156

PROFESSIONAL TRAINING OF SHIPS' PERSONNEL IN THE SYSTEM OF MILITARY PRODUCT LIFE CYCLE

E.F. Losev¹, Dr.Sc. (Military Sciences), Professor, captain 1st rank, retired, losev1947@mail.ru

I.V. Kuznetsov², Ph.D. (Military Sciences), captain 2nd rank, 89817190035@mail.ru

A.A. Bavula¹, captain 2nd rank, adjunct, bavel_@mail.ru; **I.A. Burik**¹, captain 1st rank, Head of Chair, bavel_@mail.ru

¹ "Soviet Union Fleet Admiral N.G. Kuznetsov Naval Academy" of the Navy Military Educational and Scientific Centre (Kaliningrad branch), Sovetsky Ave. 82, Kaliningrad, 236036, Russian Federation

² Military Institute of Continuing Professional Education of the VUNTS Navy "Naval Academy", Malookhtinsky Ave. 80/2, St. Petersburg, 195112, Russian Federation

Abstract. The article discusses personnel training in a military life-cycle management system based on the so-called "end-to-end life cycle contracts".

The authors propose a conceptual model of a military life-cycle management system with integration professional training of surface ships' personnel of communication departments into a common information space.

The paper presents a critical aspect of the integrated simulator of ship's communicators "Tribe-S".

The authors analyze foreign and historical experience of military professional training. They assess the quality of personnel training based on virtual environment simulation learning of navy crews during transition to end-to-end life cycle contracts.

The authors suggest that the virtual environment simulation will allow trainees to acquire unique skills in a variety of emergency situations, which sometimes is not possible to gain in a traditional learning process.

The authors also consider that the contracts based on military production life-cycle management will increase the military preparedness of a ship in general and will improve the quality of service staff professional education. The exploitation of a communication system supplied to the Navy will decrease the number of facilitators significantly. The transition to these contracts will allow more efficient use of huge funds allocated to defending our country.

Keywords: professional training, military products life cycle, virtual environment simulation learning.

References

1. Zakharov V.L., Ilyin V.A., Kushnarev A.G. Operational-tactical system of the Naval Academy. What should it be? *Oborony zakaz* [Defense order]. 2010, no. 28 (in Russ.).
2. Dovzhenko V.N., Struchkov A.M., Turovsky O.M. The use of modern information technologies in the navy personnel training system. *Morskoy sbornik*. 2009, no. 12, pp. 44–51 (in Russ.).
3. Robert I.V. *Sovremennye informatsionnye tekhnologii v obrazovanii* [Modern information technologies in education]. Moscow, IIE RAE Publ., 2010, 107 p.
4. Bakardzhieva S. Contracts of life cycle products as a motivation to modernization. *Umnoe proizvodstvo* [Intelligent Manufacturing]. 2013, no. 3, pp. 65–67 (in Russ.).
5. Sergeev V.V., Puchko E.V., Rodionov A.V. The image of future simulators naval communications. *Programmnye produkty i sistemy* [Software & Systems]. 2015, no. 1, pp. 92–99 p (in Russ.).
6. Volkova V.N., Denisov A.A. *Teoriya sistem* [The system theory]. Moscow, Vyssh. shk. Publ., 2006, 511 p.
7. Pechnikov A.N., Vetrov Yu.A. *Proektirovaniye i primeneniye kompyuternykh tekhnologiy obucheniya. Ch. 1. Kontseptsiya i primeneniye kompyuternykh tekhnologiy obucheniya* [Design and Use of Information Technologies in Education. P. 1. Concept and Use of Computer Technologies in Education]. St. Petersburg. BGТУ Publ., 2003, Book 1, 195 p.
8. *Metodicheskoe rukovodstvo po razrabotke stsennarnykh materialov dlya avtomatizirovannykh uchebnykh zanyaty s ispolzovaniem kompleksa programm instrumentalnykh sredstv "MediaTor"* [Methodical Guidance on the Development of Scenarios for Automated Training Sessions Using "MediaTor"]. St. Petersburg, RI CPS Publ., 2002.
9. Bashmakov A.I., Bashmakov I.A. *Razrabotka kompyuternykh uchebnikov i obuchayushchikh sistem* [Design of Computer Textbooks and Training Systems]. Moscow, Philin Publ., 2003, 616 p.

СОДЕРЖАНИЕ

Зубкова Т.М., Наточая Е.Н.

Проектирование интерфейса программного обеспечения с использованием элементов искусственного интеллекта 5

Пучков Е.В., Терехов С.

Разработка системы хранения ансамблей нейросетевых моделей 12

Бурдо Г.Б., Семенов Н.А.

Интеллектуальная поддержка принятия решений при диспетчировании технологических процессов в многономенклатурном машиностроении 21

Еремеев А.П., Кожухов А.А.

Реализация методов обучения с подкреплением на основе темпоральных различий и мультиагентного подхода для интеллектуальных систем реального времени 28

Шахов В.В., Юргенсон А.Н., Соколова О.Д.

Моделирование воздействия атаки Black Hole на беспроводные сети 34

Тиханычев О.В.

Прогнозирование при управлении динамическими системами 40

Тхан В.З., Берчук Д.Ю.

Системы автоматического управления объектами с запаздыванием: робастность, быстроедействие, синтез 45

Сухобоков А.А., Строгонова В.И.

Об одном подходе к реализации системы управления мастер-данными об активах 51

Кораблин Ю.П., Шипов А.А.

Построение моделей систем на базе эквивалентной характеристики формул LTL 61

Кожомбердиева Г.И., Бураков Д.П., Гарина М.И.

Использование формулы Байеса при оценивании выполнения практик модели CMMI® 67

Бобков В.А., Кудряшов А.П., Мельман С.В.

Многопроцессорная обработка в задаче пространственной реконструкции по множеству видов 75

Трусов А.Н., Иванченко П.Ю., Кацура Д.А.

Редактирование и внесение информации в XML-документы автоматизированных информационных систем 81

Батура Т.В.

Методы автоматической классификации текстов 85

Козлов П.Ю.

Методы автоматизированного анализа коротких неструктурированных текстовых документов 100

Шальнов Е.В., Конушин А.С. Использование геометрии сцены для увеличения точности детекторов	106
Хрянин Е.Л., Швецов А.Н. Программный агент определения психологического состояния обучаемого в системах дистанционного обучения	112
Сухов А.В., Решетников В.Н., Савилкин С.Б. Мониторинг частотного ресурса геостационарных спутников ретрансляторов с использованием энтропии покрытия.....	119
Агамиров Л.В., Агамиров В.Л., Вестяк В.А. Статистический анализ результатов испытаний изделий авиационной техники в условиях случайного цензурирования	124
Викторов А.С. Алгоритм детектирования объектов на фотоснимках с низким качеством изображения.....	130
Юй Чуцяо, Бессмертный И.А. Автоматический синтаксический анализ китайских предложений при ограниченном словаре	138
Осколков В.М., Варфоломеев И.А., Виноградова Л.Н., Ершов Е.В. Разработка и внедрение модели прогнозирования цветового отклонения полимерного покрытия оцинкованной полосы	143
Марголис Б.И. Программа идентификации условий теплообмена для изделий плоской формы	148
Лосев Е.Ф., Кузнецов И.В., Бавула А.А., Бурик И.А. Профессиональная подготовка личного состава корабля в системе жизненного цикла продукции военного назначения	152

Гостиница «Октябрьская»

Добро пожаловать в современный уютный отель в Твери!

- Комфортные условия проживания, приветливый персонал.
- Номера различных категорий – от «эконом» до «люкс».
- Низкие цены с включенным завтраком.
- Кафе.
- Бесплатные Wi-Fi и охраняемая автостоянка.

Адрес: г. Тверь, просп. 50 лет Октября, д. 3г
Бронирование по телефонам: 8 (4822) 44-44-94, 49-46-94
Бронирование на сайте: www.oktoberhotel.ru

C O N T E N T

Zubkova T.M., Natochaya E.N. Software interface design using elements of artificial intelligence	5
Puchkov E.V., Terekhov S. Neural network ensembles storage development	12
Burdo G.B., Semenov N.A. Intelligent decision support in process scheduling in diversified engineering	21
Eremeev A.P., Kozhukhov A.A. Implementation of reinforcement learning methods based on temporal differences and a multi-agent approach for real-time intelligent systems	28
Shakhov V.V., Yurgenson A.N., Sokolova O.D. Modelling and simulation of Black Hole attack on wireless networks	34
Tikhanychev O.V. Forecasting in dynamic system control	40
Than V.D., Berchuk D.Yu. Time delay automatic control systems: robustness, response time, synthesis	45
Sukhobokov A.A., Strogonova V.I. On an approach to construct asset master data management system	51
Korablin Yu.P., Shipov A.A. System models construction based on LTL formula equational characteristics	61
Kozhemberdieva G.I., Burakov D.P., Garina M.I. Using Bayes' theorem to estimate CMMI® practices implementation	67
Bobkov V.A., Kudryashov A.P., Melman S.V. Multiprocessing for spatial reconstruction based on multiple range-scans	75
Trusov A.N., Ivanchenko P.Yu., Katsuro D.A. Editing and entering information into XML-documents of automated information systems	81
Batura T.V. Automatic text classification methods	85
Kozlov P.Yu. Automated analysis method of short unstructured text documents	100
Shalnov E.V., Konushin A.S. Scene geometry for detector precision improvement	106
Khryanin E.L., Shvetsov A.N. A software agent to determine student's psychological state in e-learning systems	112
Sukhov A.V., Reshetnikov V.N., Savilkin S.B. Monitoring of frequency resource of geostationary repeater satellites using cover entropy	119
Agamirov L.V., Agamirov V.L., Vestyak V.A. Statistical analysis of test results of products of aeronautical engineering in terms of random censoring	124
Viktorov A.S. Object detection algorithm in low image quality photographs	130
Yu Chuqiao, Bessmertny I.A. Automatic syntactic analysis of chinese sentences by a restricted dictionary	138
Oskolkov V.M., Varfolomeev I.A., Vinogradova L.N., Ershov E.V. Development and implementation of a forecasting model for a chromatic error of galvanized strip polymeric coating	143
Margolis B.I. Program for heat conditions identification in flat products	148
Losev E.F., Kuznetsov I.V., Bavula A.A., Burik I.A. Professional training of ships' personnel in the system of military product life cycle	152

Разработки НИИ «Центрпрограммсистем»

Комплексные решения по созданию систем защиты информации для существующих и разрабатываемых автоматизированных (информационных) систем

- Системы защиты конфиденциальной информации.
- Системы защиты персональных данных до уровня К1.
- Системы защиты информации от несанкционированного доступа и иностранных технических разведок, позволяющие обрабатывать информацию, отнесенную к государственной тайне, с грифом до «совершенно секретно» включительно.
- Аппаратно-программные модули доверенной загрузки.
- Аппаратно-программные комплексы защиты информации.

Разработки ведутся строго в соответствии с требованиями российского законодательства, руководящих документов ФСТЭК России, ФСБ России, нормативных документов МО РФ и других ведомств.

Правомочность выполнения работ подтверждается действующими лицензиями ФСБ, ФСТЭК и Минобороны России.

Опыт работы НИИ «Центрпрограммсистем» в области защиты информации (информационной безопасности) – более 20 лет.

*Подробности на сайте www.szi-cps.ru
Телефон в Москве: (495) 998-60-05
Телефон в Твери: (4822) 68-06-36*

Полный комплекс работ по защите гостайны

- Аттестационные испытания на соответствие требованиям безопасности с выдачей аттестата соответствия.
- Специальные проверки для выявления технических средств перехвата информации (закладных устройств).
- Специальные исследования для выявления возможных каналов утечки информации посредством побочных электромагнитных излучений и наводок.

Преимущества разработчика:

- крупнейшая IT-компания с 45-летней историей;
- более чем 20-летний опыт работы в области защиты информации;
- собственная лаборатория, отвечающая всем нормативным требованиям;
- наличие всех необходимых сертификатов и лицензий;
- современная контрольно-измерительная аппаратура, включая сертифицированные ФСТЭК России автоматизированные комплексы.

*Подробности на сайте www.cpstver.ru
Телефон в Москве: (495) 998-59-77
Телефон в Твери: (4822) 69-06-23*

Вниманию подписчиков

Международный журнал
«Программные продукты и системы»
распространяется по подписке
в странах СНГ и за их пределами.
В Каталоге подписных изданий
Агентства «Роспечать» индекс журнала
70799

Телефон (482-2) 39-91-49

Факс (482-2) 39-91-00

www.swsys.ru

red@cps.tver.ru