

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Том 21
2015
№ 5

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г.

УЧРЕДИТЕЛЬ

Издательство "Новые технологии"

СОДЕРЖАНИЕ

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

- Захлебин И. В., Фомичев В. А.** Разработка метода семантического поиска специализированных в корпоративной базе данных по естественно-языковым запросам . . . 323
- Кулагин В. П., Цветков В. Я., Лапчинская М. П.** Идентификация морских объектов при сильном волнении водной поверхности . . . 332
- Лучкова С. В., Перемитина Т. О., Яценко И. Г.** Использование нечеткого моделирования для повышения репрезентативности информации на примере анализа характеристик нефти . . . 336

МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ

- Кухаренко Б. Г., Солнцева-Чалей М. О.** Применение моделей нелинейных динамических систем для анализа результатов кластеризации многомерных траекторий . . . 341
- Мишин И. В.** Численное моделирование однолинейных систем обслуживания . . . 346

ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ И СЕТИ

- Оцоков Ш. А.** Ускорение высокоточных вычислений за счет распараллеливания операции округления в комплексе систем счисления . . . 352
- Федюнин Р. Н.** Оценка пространственной сложности функциональных блоков АЛУ на базе однородных вычислительных структур . . . 356

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

- Черняховская Л. Р., Герасимова И. Б., Давлетбаева А. Р.** Разработка системы нечеткого логического вывода для анализа результативности процесса дистанционного обучения . . . 362

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В БИОМЕДИЦИНСКИХ СИСТЕМАХ

- Захаров М. С.** Динамика во времени вегетативных показателей на основе компьютерного спектрально-временного анализа ритмограмм сердца . . . 367

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

- Зак Ю. А.** Об одной задаче построения допустимых и оптимальных маршрутов доставки грузов . . . 373
- Панкратьев П. С., Шакиров В. А.** Двухуровневый многокритериальный анализ пунктов строительства гидроэлектростанции . . . 385

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОРГАНИЗАЦИОННЫХ И СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ СИСТЕМАХ

- Андрейчикова О. Н., Вавилов А. А., Козырев А. Н.** Исследование поведения потребителей на рынке программного обеспечения . . . 391

Главный редактор:

СТЕМПКОВСКИЙ А. Л.,
акад. РАН, д. т. н., проф.

Зам. главного редактора:

ИВАННИКОВ А. Д., д. т. н., проф.
ФИЛИМОНОВ Н. Б., д. т. н., с.н.с.

Редакционный совет:

БЫЧКОВ И. В., акад. РАН, д. т. н.

ЖУРАВЛЕВ Ю. И.,
акад. РАН, д. ф.-м. н., проф.

КУЛЕШОВ А. П.,
акад. РАН, д. т. н., проф.

ПОПКОВ Ю. С.,
чл.-корр. РАН, д. т. н., проф.

РУСАКОВ С. Г.,
чл.-корр. РАН, д. т. н., проф.

РЯБОВ Г. Г.,
чл.-корр. РАН, д. т. н., проф.

СОЙФЕР В. А.,
чл.-корр. РАН, д. т. н., проф.

СОКОЛОВ И. А., акад.
РАН, д. т. н., проф.

СУЕТИН Н. В., д. ф.-м. н., проф.

ЧАПЫГИН Ю. А.,
чл.-корр. РАН, д. т. н., проф.

ШАХНОВ В. А.,
чл.-корр. РАН, д. т. н., проф.

ШОКИН Ю. И.,
акад. РАН, д. т. н., проф.

ЮСУПОВ Р. М.,
чл.-корр. РАН, д. т. н., проф.

Редакционная коллегия:

АВДОШИН С. М., к. т. н., доц.

АНТОНОВ Б. И.

БАРСКИЙ А. Б., д. т. н., проф.

ВАСЕНИН В. А., д. ф.-м. н., проф.

ВИШНЕКОВ А. В., д. т. н., проф.

ГАЛУШКИН А. И., д. т. н., проф.

ДИМИТРИЕНКО Ю. И., д. ф.-м. н., проф.

ДОМРАЧЕВ В. Г., д. т. н., проф.

ЗАГИДУЛЛИН Р. Ш., к. т. н., доц.

ЗАРУБИН В. С., д. т. н., проф.

КАРПЕНКО А. П., д. ф.-м. н., проф.

КОЛИН К. К., д. т. н., проф.

КУЛАГИН В. П., д. т. н., проф.

КУРЕЙЧИК В. М., д. т. н., проф.

КУХАРЕНКО Б. Г., к. ф.-м. н., доц.

ЛЬВОВИЧ Я. Е., д. т. н., проф.

МИХАЙЛОВ Б. М., д. т. н., проф.

НЕЧАЕВ В. В., к. т. н., проф.

ПОЛЕЩУК О. М., д. т. н., проф.

СОКОЛОВ Б. В., д. т. н., проф.

УСКОВ В. Л., к. т. н. (США)

ФОМИЧЕВ В. А., д. т. н., проф.

ЧЕРМОШЕНЦЕВ С. Ф., д. т. н., проф.

ШИЛОВ В. В., к. т. н., доц.

Редакция:

БЕЗМЕНОВА М. Ю.

ГРИГОРИН-РЯБОВА Е. В.

ЛЫСЕНКО А. В.

ЧУГУНОВА А. В.

Информация о журнале доступна по сети Internet по адресу <http://novtex.ru/IT>.

Журнал включен в систему Российского индекса научного цитирования.

Журнал входит в Перечень научных журналов, в которых по рекомендации ВАК РФ должны быть опубликованы научные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

INFORMATION TECHNOLOGIES

INFORMACIONNYYE TEHNOLOGII

Vol. 21
2015
No. 5

THEORETICAL AND APPLIED SCIENTIFIC AND TECHNICAL JOURNAL

Published since November 1995

ISSN 1684-6400

CONTENTS

INTELLIGENT SYSTEMS AND TECHNOLOGIES

- Zakhlebin I. V., Fomichov V. A.** Development of a Method for Semantic Search of Specialists in Corporate Databases using Natural Language Queries 323
- Kulagin V. P., Tsvetkov V. Ya., Lapchinskaya M. P.** Identification of Ships in Heavy Seas 332
- Luchkova S. V., Peremitina T. O., Yashchenko I. G.** Usage Fuzzy Modeling for Raising a Sample Representative on the Example the Analysis of Characteristics Oils 336

MODELING AND OPTIMIZATION

- Kukhareno B. G., Solntseva-Chalei M. O.** Applying Nonlinear Dynamical System Models for Analysis of Multidimensional Trajectory Clustering Results 341
- Mishin I. V.** Numerical Modeling of Single-Server Systems 346

COMPUTING SYSTEMS AND NETWORKS

- Otsokov Sh. A.** Acceleration of High-Precision Computation Based on Parallelization of Group Number Systems 352
- Fedjunin R. N.** The Evaluation of Hardware Complexity Functional Part of ALU Based Systolic Structure 356

INFORMATION TECHNOLOGIES IN EDUCATION

- Chernyakhovskaya L. R., Gerasimova I. B., Davletbaeva A. R.** Development of Fuzzy Inference System for Distance Learning Process Analysis 362

INFORMATION TECHNOLOGIES IN BIOMEDICAL SYSTEMS

- Zakharov M. S.** Dynamics in Time Vegetative Indices Based on Computer Spectral Analysis of Heart Rhythmograms 367

APPLIED INFORMATION TECHNOLOGIES

- Zack Yu. A.** A Problem of Constructing Admissible and Optimal Routes of Delivery 373
- Pankratiev P. S., Shakirov V. A.** Two-Level Multiattribute Analysis of Hydroelectric Station Sites 385

INFORMATION TECHNOLOGIES IN THE ORGANIZATIONAL AND SOCIO-ECONOMIC SYSTEMS

- Andreichicova O. N., Vavilov A. A., Kozyrev A. N.** The Investigation of Consumer Behavior at the Software Market 391

Editor-in-Chief:

Stempkovsky A. L., Member of RAS,
Dr. Sci. (Tech.), Prof.

Deputy Editor-in-Chief:

Ivannikov A. D., Dr. Sci. (Tech.), Prof.
Filimonov N. B., Dr. Sci. (Tech.), Prof.

Chairman:

Bychkov I. V., Member of RAS,
Dr. Sci. (Tech.), Prof.
Zhuravljov Yu. I., Member of RAS,
Dr. Sci. (Phys.-Math.), Prof.
Kuleshov A. P., Member of RAS,
Dr. Sci. (Tech.), Prof.
Popkov Yu. S., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Rusakov S. G., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Ryabov G. G., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Soifer V. A., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Sokolov I. A., Member of RAS,
Dr. Sci. (Phys.-Math.), Prof.
Suetin N. V.,
Dr. Sci. (Phys.-Math.), Prof.
Chaplygin Yu. A., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Shakhnov V. A., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Shokin Yu. I., Member of RAS,
Dr. Sci. (Tech.), Prof.
Yusupov R. M., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.

Editorial Board Members:

Avdoshin S. M., Cand. Sci. (Tech.), Ass. Prof.
Antonov B. I.
Barsky A. B., Dr. Sci. (Tech.), Prof.
Vasenin V. A., Dr. Sci. (Phys.-Math.), Prof.
Vishnekov A. V., Dr. Sci. (Tech.), Prof.
Galushkin A. I., Dr. Sci. (Tech.), Prof.
Dimitrienko Yu. I., Dr. Sci. (Phys.-Math.), Prof.
Domrachev V. G., Dr. Sci. (Tech.), Prof.
Zagidullin R. Sh., Cand. Sci. (Tech.), Ass. Prof.
Zarubin V. S., Dr. Sci. (Tech.), Prof.
Karpenko A. P., Dr. Sci. (Phys.-Math.), Prof.
Kolin K. K., Dr. Sci. (Tech.)
Kulagin V. P., Dr. Sci. (Tech.), Prof.
Kureichik V. M., Dr. Sci. (Tech.), Prof.
Kukhareno B. G., Cand. Sci. (Phys.-Math.)
Ljvovich Ya. E., Dr. Sci. (Tech.), Prof.
Mikhailov B. M., Dr. Sci. (Tech.), Prof.
Nechaev V. V., Cand. Sci. (Tech.), Ass. Prof.
Poleschuk O. M., Dr. Sci. (Tech.), Prof.
Sokolov B. V., Dr. Sci. (Tech.)
Uskov V. L. (USA), Dr. Sci. (Tech.)
Fomichev V. A., Dr. Sci. (Tech.), Prof.
Chermoshentsev S. F., Dr. Sci. (Tech.), Prof.
Shilov V. V., Cand. Sci. (Tech.), Ass. Prof.

Editors:

Bezmenova M. Yu.
Grigorin-Ryabova E. V.
Lysenko A. V.
Chugunova A. V.

Complete Internet version of the journal at site: <http://novtex.ru/IT>.

According to the decision of the Higher Certifying Commission of the Ministry of Education of Russian Federation, the journal is inscribed in "The List of the Leading Scientific Journals and Editions wherein Main Scientific Results of Theses for Doctor's or Candidate's Degrees Should Be Published"

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ INTELLIGENT SYSTEMS AND TECHNOLOGIES

УДК 004.512:004.89

И. В. Захлебин, студент магистратуры, **В. А. Фомичев**, д-р техн. наук, проф., e-mail: vfomichov@hse.ru, Национальный исследовательский университет "Высшая школа экономики", Москва

Разработка метода семантического поиска специалистов в корпоративной базе данных по естественно-языковым запросам

Обосновывается актуальность проблемы создания компьютерных систем семантического поиска специалистов по естественно-языковым описаниям их компетенций. Описывается разработанный и программно реализованный на языке Python метод проектирования прикладных компьютерных систем этого класса. Важной отличительной чертой метода является использование класса СК-языков (стандартных концептуальных языков), предлагаемого теорией К-представлений, для формализации семантики естественно-языковых запросов и представления семантики слов и коротких словосочетаний в лингвистической базе данных.

Ключевые слова: семантический поиск специалистов, естественно-языковой запрос, семантическое представление, теория К-представлений, СК-язык, концептуальный базис, лингвистическая база данных

Введение

Сегодня компании вынуждены осуществлять непрерывный поиск специалистов, например, для набора персонала на открывшиеся вакансии, образования команд, устранения нештатных ситуаций и организации внутренней мобильности [1]. Ускорение поиска компетентных сотрудников дает им конкурентное преимущество: компания IBM, внедрив в начале 2000-х годов собственную систему поиска специалистов, смогла к 2006 г. сэкономить более \$500 млн [2].

Как правило, компании имеют базы данных (БД), содержащие профили тысяч соискателей и собственных сотрудников. Эффективная обработка такого объема информации выходит за пределы человеческих возможностей. В связи с этим возникает потребность в разработке интеллектуальных систем, позволяющих автоматизировать часть задач по отбору кандидатов.

Заметим, что определение компетенций специалистов по неструктурированной текстовой информации (например, по резюме или рабочим документам) является отдельной сложной задачей. Для ее решения предложен ряд эффективных подходов, например, в работах [3–5]. Поэтому мы будем рассматривать проблему поиска специалистов в реалиях больших компаний, используя структурированную информацию о них из имеющейся корпоративной БД и предполагая, что она могла быть получена не только путем целенаправленного сбора, но и автоматически.

Отечественные системы поиска специалистов прежде всего представлены Интернет-ресурсами,

агрегирующими резюме и предоставляющими поиск по ним. Пятью наиболее популярными в России ресурсами этой категории являются "SuperJob", "HeadHunter", "JOB.RU", "Работа.RU" и "НГС.Работа"¹. Данные системы способны эффективно работать только с критериями отбора, заданными в структурированном виде, как правило, при заполнении специальных форм. Возможности же использования в них запросов на естественном языке (ЕЯ-запросов) ограничены простым поиском по ключевым словам. В данной области, насколько известно авторам, более совершенных методов обработки запросов для русского языка на практике не реализовано.

Но стоит заметить, что ЕЯ-запросы являются наиболее удобным для пользователя системы способом выражения своих мыслей. Исследование [6] показывает, что пользователи (на примере выборки из 48 человек) быстрее и с более высоким качеством справлялись с задачей поиска, используя системы с естественно-языковым интерфейсом (ЕЯ-интерфейсом). Сравнение проводилось с системами, где запросы составлялись графически и с помощью контекстных меню. ЕЯ-интерфейсы испытывают меньшие трудности при масштабировании и адаптации к новым форматам данных по сравнению с аналогами.

На первый взгляд, для поиска сотрудников по текстовым запросам приемлемо использовать традиционные методы поиска по коллекциям текстов вплоть до применения готовых технологий, например Google Custom Search Engine². Но такие сис-

¹Измерено по статистике посещаемости ресурсов на сервисе LiveInternet по состоянию на 1 ноября 2014 г.

²<http://www.google.com/cse/>

темы в силу своего общего назначения не полностью учитывают при поиске имеющуюся структуру информации. Поэтому они склонны выводить результаты, в которых найдены совпадения в неправильных полях, или же результаты, не соответствующие всем критериям.

С начала 2000-х годов можно наблюдать значительный прогресс в разработке лингвистических процессоров (ЛП) прикладных компьютерных систем. В частности, в разных странах и разных исследовательских центрах разрабатывались:

- ЕЯ-интерфейсы рекомендательных систем [7, 8];
- ЕЯ-интерфейсы вопросно-ответных систем для взаимодействия с онтологиями, разработанными в рамках или под влиянием проекта Семантический Веб [9, 10];
- преобразователи описаний фрагментов знаний на ЕЯ в наборы выражений языка проектирования онтологий OWL [11];
- ЛП для семантической обработки текстов по биологии и медицине как одно из центральных направлений биоинформатики [12].

Таким образом, в контексте современного состояния исследований по разработке лингвистических информационных технологий представляется актуальной проблема создания компьютерных систем семантического поиска специалистов по естественно-языковым описаниям их компетенций. Целью данной статьи, продолжающей линию работы [13], является разработка метода создания таких систем на основе применения теории К-представлений (концептуальных представлений) для формализации семантики запросов и отображения семантики слов и коротких словосочетаний в лингвистической базе данных. Важным рассматриваемым требованием при поиске специалистов является использование системой структуры имеющейся корпоративной БД.

Для достижения поставленной цели в статье решаются следующие задачи:

- выбирается способ формального представления структурированных значений ЕЯ-текстов;
- разрабатывается метод семантического поиска специалистов по ЕЯ-запросам;
- программно реализуется система семантического поиска, работающая с корпоративной БД.

Выбор подхода к формальному представлению значений текстов

С начала 1980-х годов исследователи из разных стран искали более удобные и широко применимые методы формального описания семантической структуры ЕЯ-текстов по сравнению с языками логики предикатов первого порядка (ЛППП) и более математически развитые подходы (более тонко структурированные) по сравнению с теорией семантических сетей. В этот период были, в частности, разработаны теория представления дискурсов (Discourse Representation Theory), теория концептуаль-

ных графов (Theory of Conceptual Graphs), эпизодическая логика (Episodic Logic), теория расширенных семантических сетей, компьютерная семантика русского языка и теория неоднородных семантических сетей.

Среди всех указанных подходов своим замыслом и результатами выделяется теория К-представлений (концептуальных представлений). Эта теория, первоначально называвшаяся теорией К-исчислений и К-языков, представлена в большой серии публикаций на русском и английском языках, в том числе в работах [14—21]. В отличие от всех перечисленных выше подходов, цель разработки теории К-представлений (ТКП) заключалась в поиске системы операций на концептуальных структурах, позволяющих шаг за шагом строить формальное представление структурированного значения, или семантическое представление (СП), произвольно сложных предложений и связанных текстов (или дискурсов) на русском, английском, немецком, французском и других языках; совокупность таких языков в лингвистике называется естественным языком (ЕЯ).

Базовая модель ТКП описывает систему, состоящую из 10 частичных операций на концептуальных структурах. Эта модель задает новый класс формальных языков — класс СК-языков (стандартных концептуальных языков) [14—17, 20, 21]. Общим преимуществом СК-языков по сравнению с перечисленными подходами к формализации семантики ЕЯ-текстов является уникальный набор новых выразительных механизмов, открывающих возможности моделирования семантической структуры словосочетаний и текстов следующих видов (при сохранении выразительных средств ЛППП):

- инфинитивных и герундиальных конструкций (выражающих действия, цели, обязательства и т. д.);
- фраз со сложными придаточными предложениями, в том числе с прямой и косвенной речью, с придаточными цели;
- произвольно сложных составных обозначений множеств и понятий;
- связанных текстов (дискурсов) со ссылками на смысл предыдущих или последующих фраз и более крупных частей текста;
- обозначений упорядоченных наборов объектов [14—17, 20, 21].

ТКП включает широко применимую математическую модель лингвистической базы данных (ЛБД) и несколько сильно структурированных алгоритмов семантико-синтаксического анализа ЕЯ-текстов, использующих эту модель ЛБД [17—19, 21]. Модель и алгоритмы применены при разработке рекомендательной системы с ЕЯ-интерфейсом [22], информационно-поисковой системы, реализующей семантический поиск по ЕЯ-запросам пользователей [23] и при проектировании системы управления файлами с ЕЯ-интерфейсом [24]. Теория К-представ-

лений включает оригинальную концепцию преобразования Веба в Мультилингвистический Семантический Веб [21, 25].

На основании перечисленных выше аргументов в качестве базового подхода к формальному описанию семантической структуры ЕЯ-текстов в проводимом исследовании была выбрана теория К-представлений.

Для построения СП ЕЯ-текстов из первичных информационных единиц и нескольких служебных символов теория К-представлений вводит набор некоторых правил P[0]—P[10]. Каждое из правил задает частичную операцию на множестве концептуальных структур. По гипотезе В. А. Фомичева [14—17, 20, 21] данный набор правил позволяет строить СП сколь угодно сложных ЕЯ-текстов. Перечислим здесь лишь те из них, которые непосредственно используются в работе:

- правило P[0] задает по концептуальному базису множество простейших формул (простейшие и сложные формулы основного класса называются К-цепочками);
- по правилу P[1] образуются К-цепочки вида *int.qtr concept*, где *int.qtr* — интенциональный квантор (семантическая единица, соответствующая словам "некоторый", "произвольный", "все" и т. д.), *concept* — простое или составное обозначение понятия;
- правила P[2] и P[3] позволяют соответственно строить К-цепочки видов $f(a_1, \dots, a_n)$ и $(b_1 \equiv b_2)$, где $n \geq 1$, f — n -арный функциональный символ, $a_1, \dots, a_n, b_1, b_2$ — К-цепочки;
- с помощью P[4] строятся К-цепочки вида $r(a_1, \dots, a_n)$, где $n \geq 1$, r — n -арный реляционный символ;
- правило P[7] использует логическую связку \wedge (конъюнкция) или \vee (дизъюнкция) для построения К-цепочек вида $(a_1 \wedge a_2 \wedge \dots \wedge a_n)$, где $n > 1$;
- по правилу P[8] строятся К-цепочки вида $c * (r_1, b_1) \dots (r_n, b_n)$, где c — базовое понятие, описывающее объект, а r_1, \dots, r_n — бинарные отношения и b_1, \dots, b_n — соответствующие значения вторых атрибутов;
- правило P[10] нужно для формирования обозначений упорядоченных наборов объектов — К-цепочек вида $\langle a_1, \dots, a_n \rangle$, где для $i = 1, \dots, n$ элемент a_i — более простая К-цепочка.

Структура разработанной поисковой системы

Для реализации предлагаемого подхода к семантическому поиску специалистов разработана система-прототип ExpSearch (Experts Search). Ее структура изображена на рисунке. Она похожа на структуру системы, предложенной в работе [13], но отличается составом лингвистической базы данных (ЛБД) и выделением отдельного компонента для интерпретации запросов вместо подсистемы поиска по ключевым словам. Предполагается, что перед началом использования системы имеется некоторая



Структура системы ExpSearch

локально доступная БД с загруженной в нее информацией о специалистах. Также предполагается, что эта БД обновляется с некоторой периодичностью, в связи с чем в структуру системы включена подсистема сбора информации. В частности, в работе используется БД, собранная на основе данных из Интернета с помощью указанной подсистемы.

Человек взаимодействует с системой через интерфейс пользователя, его основа — поле для ввода запроса. Семантико-синтаксический анализатор служит для построения СП запросов пользователя. В работе он опирается на ЛБД, служащую для связи слов, входящих в запросы, с соответствующими им концептуальными единицами и для нахождения семантико-синтаксических связей между элементами текста. На выходе анализатор выдает СП запроса в установленном формате.

Интерпретатор запросов по построенному СП пользовательского запроса конструирует запрос к БД, записанный на стандартном языке, например SQL или JSON. Модель интерпретации запросов показывает, как транслировать конструкции используемого СК-языка в формальные конструкции целевого языка для составления синтаксически корректных запросов.

• Используемые данные

В качестве источника фактической информации о наборе специалистов в данной работе было решено использовать информацию о сотрудниках Национального исследовательского университета "Высшая школа экономики" (НИУ ВШЭ), размещенную на официальном сайте вуза³. Данный выбор обусловлен нахождением всех материалов в открытом доступе, формальностью описания компетенций и большим объемом выборки специалистов.

Профили сотрудников ВШЭ содержат следующие разделы (звездочкой отмечены обязательные):

- 1) фамилия, имя и отчество*;
- 2) подразделение и занимаемая должность*;
- 3) высшее образование (уровень, вуз, факультет, специальность и год окончания);
- 4) опыт работы (периоды трудоустройства и названия организаций);

³<http://www.hse.ru/org/persons/>

5) проводимые курсы в текущем академическом году (номер курса, факультет, специальность, название);

6) профессиональные интересы;

7) публикации (авторы, название, сведения об издании).

Приведенный набор данных содержит информацию как о достижениях сотрудников, так и о сферах их интересов. Он в значительной степени отражает компетенции специалистов и может быть использован для поиска по ним.

Используемая системой информация хранится в БД и имеет определенный формат. Так, сведения об образовании представлены в виде пятерок <уровень образования, институт, факультет, специальность, год выпуска>.

• Структура концептуального базиса

В теории К-представлений на первом этапе определения класса СК-языков вводится понятие концептуального базиса (к.б.). Задание произвольного к.б. равносильно заданию длинного набора формальных объектов: множеств символов, выделенных элементов множеств, некоторой функции и двух бинарных отношений на одном из множеств. Такой набор задает, в частности, множество X рассматриваемых первичных информационных единиц и сведения, связанные с элементами множества X и необходимые для построения СП ЕЯ-текстов. В множестве X выделяется конечное подмножество St , элементы которого называются сортами и интерпретируются как наиболее общие понятия, рассматриваемые в выбранной области (физический объект, организация и т. д.) [15–17].

В проведенном исследовании для описания сущностей из рассматриваемой области введена специально адаптированная для нее система сортов. Например, введены сорта, относящиеся к работе и образованию специалиста: *длительность, организация, должность* и *год-выпуска*. Заметим, что введенные сорта можно упорядочить с помощью отношения общности *Gen* (частичного порядка), задающего иерархию на множестве сортов St . Например, сорт *время* является более общим, чем сорта *длительность* и *год-выпуска*, что дает пары (*время, длительность*) и (*время, год-выпуска*), входящие в отношение *Gen*.

• Формат семантических представлений запросов

Заметим, что традиционные системы поиска специалистов полагаются на заполнение полей форм для задания соответствующих критериев отбора. Данный подход является обоснованным, особенно с учетом того, что сами специалисты привыкли структурировать информацию о себе по не связанным друг с другом категориям, содержащим записи в единообразном формате. Соответственно, базы данных имеют схожую структуру, и в нашей системе логично использовать формат запросов, который бы позволял в текстовом виде одновременно задавать множество критериев отбора.

Для представления смысла запросов наиболее подходят К-цепочки вида $ref\ c * (r_1, b_1) \dots (r_n, b_n)$, построенные с применением правил P[8] и P[1]. Здесь *ref*-квантор референтности (в примерах — единица *нек*, от слова "некоторый"), *c* — базовое понятие, описывающее объект (в нашем случае — *чел*, "человек"), r_1, \dots, r_n и b_1, \dots, b_n — бинарные отношения и значения их вторых атрибутов [15–17].

Заметим, что критерии отбора могут быть составными. Например, для описания характеристик образования требуется определить одно из нескольких названий вуза, год выпуска и специальность. Для этого будем использовать составные значения ($b_i, i = 1, \dots, n$), построенные по правилу P[10]. Это К-цепочки вида $\langle a_1, \dots, a_n \rangle$, где a_1, \dots, a_n — отдельные более простые К-цепочки. Для каждого типа отношений последовательность и число элементов определяются схемой информации, содержащейся в БД. Если в отношении определено меньше критериев отбора, чем предусмотрено, на недостающие места ставятся нулевые элементы *nil*. Если критерий предполагает выбор из нескольких вариантов, для их комбинации используем правило P[7], определяющее операции конъюнкции и дизъюнкции ("^" и "v").

Предложена также следующая модификация языка СП. Для представления выражений вида " ≥ 2 года" расширен язык СП входных запросов путем включения в него конструкций вида "символ бинарного отношения + число + единица измерения".

Основываясь на приведенном формате запросов, введем набор отношений для задания критериев отбора специалистов, а также рассмотрим простейшие примеры запросов и их СП:

• область профессиональных интересов — (*Компетенция, область-знания*).

Пример 1: "человек, интересующийся теорией коллективного выбора".

СП 1: *нек чел * (Компетенция, теория-коллективного-выбора)*.

Пример 2: "специалист по встраиваемым системам".

СП2: *нек чел * (Компетенция, встраиваемые-системы)*;

• подразделение и должность — (*Работа, <организация, должность>*)

Пример: "профессор кафедры менеджмента".

СП: *нек чел * (Работа, <кафедра-менеджмента, профессор>)*;

• опыт работы — (*Опыт-работы, <время, организация, должность>*)

Пример: "специалист, работавший в Microsoft не менее 2 лет".

СП: *нек чел * (Опыт-работы, < ≥ 2 /год, Microsoft, nil>)*;

• образование — (*Оконч.уч.зав., <уровень-образования, институт, факультет, специальность, год-выпуска>*).

Пример: "выпускник МГУ по специальности "мировая экономика".

СП: *нек чел * (Оконч.уч.зав., <nil, МГУ, nil, мировая экономика, nil>);*

• ведение учебных курсов — (*Преподавание, <дисциплина, организация, номер-курса>*).

Пример: "преподаватель дискретной математики".

СП: *нек чел * (Преподавание, <дискретная математика, nil, nil>).*

Такой подход к конструированию запросов позволяет составлять сложные и выразительные структуры. Например, рассмотрим возможный запрос к системе: "выпускник МГУ или МИЭМ, или МИФИ, имеющий не менее чем двухлетний опыт работы программистом". Тогда возможное СП этого запроса: *нек чел * (Оконч.уч.зав., <nil, (МГУ ∨ МИЭМ ∨ МИФИ), nil, nil, nil>) (Опыт-работы, <≥2/год, nil, программист>).*

Алгоритм обработки запросов

Для того чтобы стало возможным представлять структурированные значения текстов в выбранной ранее форме (в виде К-представлений), требуется разработать алгоритм, позволяющий по фрагменту текста на ЕЯ автоматически генерировать соответствующее ему СП, а затем интерпретировать его и осуществить поиск с выделенными критериями отбора.

В книгах [17, 21] предложены алгоритмы такого рода SemSyn и SemSynt1, реализующие метод преобразования ЕЯ-текстов в СП, изложенный в работах [18, 19]. Но эти алгоритмы не устанавливают предпочтительный формат СП, без чего чрезвычайно сложно строить запросы к БД. В алгоритмах SemSyn и SemSynt1 также не предусмотрена и обработка текстов с ошибками.

В данной работе предлагается алгоритм, являющийся модификацией алгоритма SemSynt1 и ориентированный на применение в разработанной системе поиска специалистов. В частности, предложен метод выявления составных единиц в текстах. Алгоритм позволяет строить СП установленного вида без формирования матричного семантико-синтаксического представления текста. Работа алгоритма включает следующие этапы.

1. Предварительная обработка текста. Поступивший на вход алгоритма текст разбивается на токены — последовательности символов, соответствующие элементам текста: словам и разделителям. Пример: "человек, работавший в Microsoft не менее 2 лет" → ("человек", ";", "работавший", "в", "Microsoft", "не", "менее", "2", "лет").

Затем для токенов, представляющих собой слова, выполняется морфологический анализ. По информации, содержащейся в окончаниях слов, с помощью грамматического словаря определяются части речи, базовые формы слов и их морфологические характеристики.

В данной работе для обозначения морфологических признаков применяется нотация, исполь-

зуемая в Открытом корпусе русского языка — OpenCorpora⁴. Следует заметить, что обозначения, используемые для описания различных признаков, не совпадают друг с другом. Это позволяет представлять морфологические характеристики слов в виде неупорядоченных списков, что отличается от морфологического представления, описанного в работах [17, 21].

Таким образом, морфологический анализ представляет собой отображение *слово* → (*lec, Morph*), где *lec* — начальная форма слова (записанная прописными буквами), а *Morph* — список обозначений его морфологических характеристик.

2. Проекция компонентов словаря составных терминов. Для сопоставления слов ЕЯ, переведенных в начальную форму, с единицами семантического уровня удобно использовать лексико-семантический словарь (ЛСС), предложенный в работах [17, 21]. Он задает отображение (*lec, pt*) → (*sem, st₁, ..., st_k*). Начальной форме, или лексеме, *lec* и части речи *pt* слова ставится в соответствие некоторая семантическая единица *sem*, ассоциированная с набором сортов *st₁, ..., st_k*. т. е. имеющая тип *st₁ * ... * st_k*.

При таком подходе словам с ошибками единицы семантического уровня не будут сопоставлены вообще. Для замены ЛСС предлагается словарь составных терминов *Terms*, представляющий собой его обобщение. Он является массивом записей вида (*Lecs, sem, st*), компоненты которых интерпретируются следующим образом: *Lecs* — набор лексем слов, соответствующих одной единице информационного уровня (пример: единице *научно-учебная лаборатория-макрэкономического-анализа* соответствует набор {"научно-учебный", "лаборатория", "макрэкономический", "анализ"}); *sem* — К-цепочка (возможно, составная) со значением рассматриваемого словосочетания (пример: словосочетанию "старший преподаватель" сопоставляется цепочка *нек чел * (Работа, <nil, старший преподаватель>)*), так как это слово определяет должность обсуждаемого человека. Наконец, *st* — это обозначение сорта единицы *sem* (пример: единица *НИУ-ВШЭ* имеет сорт *организация*). Несколько возможных записей из ЛСС приведены в таблице. В качестве ин-

Пример записей словаря составных терминов

<i>Lecs</i>	<i>sem</i>	<i>st</i>
специалист профессор	<i>нек чел нек чел * (Работа, <nil, профессор>)</i>	<i>интел.система интел.система</i>
выпускник	<i>нек чел * (Оконч.уч.зав., <nil, nil, nil, nil>)</i>	<i>интел.система</i>
бухгалтерия высший, школа, менеджмент	<i>бухгалтерия высшая-школа- менеджмента</i>	<i>организация организация</i>
наука, о, данные дискретный, математика	<i>наука-о-данных дискретная-математика</i>	<i>специальность дисциплина</i>

⁴<http://opencorpora.org/dict.php?act=gram>

дикатора отсутствия значения используется цепочка *nil*.

Алгоритм применения словаря составных терминов можно представить следующим образом. В качестве предварительного шага строится отображение (хэш-таблица) *Map*, сопоставляющее лексемам номера записей словаря составных терминов *Terms*, в которых употребляются эти лексемы. Формально, *Map: lec* → $\{i|lec$ — компонент набора *Lecs*, являющегося *i*-й записью словаря *Terms*).

При разборе запроса каждому слову с помощью *Map* сопоставляется набор номеров релевантных ему записей. Если лексема *lec* не входит в *Map*, выбираются все элементы из области определения *Map*, расстояние Левенштейна до которых меньше некоторой величины λ , и операция проводится для них.

При проходе слева направо алгоритм комбинирует номера записей, к которым могут относиться слова запроса, и последовательно выделяет наибольшие по объему подмножества слов, имеющих одну общую запись словаря. Каждому слову из подмножества сопоставляется семантическая единица *sem* и сорт *st* из соответствующей записи. Работа алгоритма прекращается, когда всем словам поставлены в соответствие семантические единицы.

Предложенный алгоритм применения словаря составных терминов позволяет правильно обрабатывать широкий диапазон случаев искаженного написания по сравнению с записями словаря, например, опечатки (в том числе "слипания" и неправильные разбиения слов), употребление буквы "ё", а также пропуск слов или использование вместо них сокращений.

Отметим еще одну особенность алгоритма. Если после данного шага осталось много слов, не сопоставленных с семантическими единицами, с высокой степенью уверенности можно сказать, что данный запрос не будет правильно разобран. Это может происходить не только из-за неполноты словаря составных терминов или несовершенства метода разбора, но прежде всего из-за ввода запросов, не имеющих отношения к поиску специалистов.

3. Проекция словаря предложных фреймов. После выделения семантических единиц на основе слов запроса установим связи между ними для последующего построения СП. Для этого мы используем информацию о семантико-синтаксических связях между словами запроса.

В работах [17, 21] предложены словарь глагольно-предложных фреймов и словарь предложных семантико-синтаксических фреймов. Они задают связи в сочетаниях вида "глагольная форма—предлог—существительное" и "существит. 1—существит. 2" соответственно. Для того чтобы унифицировать процесс их применения, в данной работе предлагается обобщенный словарь предложных фреймов. Его записи имеют вид $(st_1, Morph_1, prep, st_2, Morph_2)$, где *prep* — необходимый предлог между словами (может быть пустым); st_1, st_2 — обозначения сортов, кото-

рые можно связать с первым и вторым словом в лингвистически правильном словосочетании " $word_1 + prep + word_2$ " соответственно; *Morph₁*, *Morph₂* — обозначения характеристик, которые должны иметь первое и второе слово соответственно (части речи, формы глагола — время/лицо/число, формы существительного — число/падеж и т. д.).

Если пара слов запроса удовлетворяет всем условиям какой-то записи словаря, считается, что между ними существует связь, и второе слово зависит от первого. Поэтому для установления связей между словами выполним следующую процедуру. Будем последовательно брать упорядоченные пары слов из запроса, их число равно $n(n - 1)$, где *n* — длина запроса в словах за вычетом знаков пунктуации и слов-связок (союзы, предлоги и т. п.). Для каждой пары проверим входящие в нее слова на соответствие условиям каждой из записей словаря. Если найдено совпадение, считается, что между этими словами имеется смысловая связь. Пара запоминается как связанная, и дальнейшая сверка по словарю для нее прекращается. Таким образом, информация о связях слов представляется асимметричным бинарным отношением $Rels \subset W \times W$ на множестве слов *W*, которое передается на следующий шаг алгоритма.

4. Построение семантического представления.

На данном шаге используем связи между отдельными словами, выделенными на предыдущих этапах, чтобы определить связи между семантическими единицами и составить итоговое семантическое представление запроса. Заметим, что для этого слова запроса, представляющие одну и ту же семантическую единицу *sem*, не должны быть обязательно связаны друг с другом отношением *Rels*, так как они уже объединены своим значением. То есть нас интересуют только связи между словами, представляющими разные единицы.

Рассмотрим все подобные связи. Согласно направлению каждой связи будем называть две рассматриваемых единицы главной и зависимой. Если в главной единице есть пустое поле (на месте аргумента стоит ключевое слово *nil*) и при этом сорт зависимой единицы совпадает с требованием к сорту этого поля (или соответствующие сорта входят в отношение *Gen*), зависимая единица подставляется в пустое поле главной. В итоге это дает на выходе финальное СП.

Заметим, что система до этого шага не использовала информацию о содержащихся в тексте пунктуационных символах. Если на данном этапе две и более семантические единицы сопоставляются с одним и тем же полем другой единицы, то по стоящим запятым и использованию союзов ("и" или "или") определяется, какую логическую связку использовать в формуле (\wedge или \vee). Это делает систему более устойчивой к некорректному вводу, так как пользователь мог пропустить символ (например, запятую между существительным и стоящим после

него зависимым причастием) или добавить лиш- ний символ.

5. Составление запроса к базе данных. На заклю- чительном этапе по построенному СП ЕЯ-запроса пользователя создается запрос к БД. Выбранная нами структура семантических представлений по- зволяет очень удобно преобразовывать их в запросы к хранилищу структурированной информации, на- пример, на языке SQL. Если в СП используются функции *Работа*, *Оконч.уч.зав.*, *Опыт-работы* или *Преподавание*, их аргументы рассматриваются как критерии, по которым отбираются специалисты. Так, при использовании одной функции *Работа* будут выбраны только специалисты с совпадающими аргументами типа *место-работы* и *должность*, если на месте одного из них в запросе не стоит ключевое слово *nil*. Это достигается тем, что в шаблон запроса подставляются условия в зависимости от содержа- ния функций.

В качестве примера покажем, как данные кон- струкции могут быть спроецированы в язык SQL. Рассмотрим функцию *Работа*. Она может задавать два критерия отбора — подразделение и должность. Тогда для условного запроса с общей частью "select person.id from position, department where person.id= department.person_id and ..." вводятся два условных выражения: "position.department=..." и "position.po- sition=...". Они заполняются соответствующими значениями функции, подставляются в секцию "where" запроса и связываются условием "and", если условий больше одного.

При использовании аргументов вида " ≥ 2 года" и подобных для соответствующих полей БД о спе- циалистах осуществляется проверка на соответст- вие указанному логическому условию. Тогда фор- мат условий немного изменяется, например, для опыта работы: "experience.duration ≥ 2 ".

Аналогичным образом обрабатывается СП, если запрос нужно сформировать в формате JSON (для No-SQL хранилища). Отличие заключается в фор- мате представления. Так, основа запроса — пустой словарь "{}", куда добавляются элементы с крите- риями. Например, для отношения *Работа*: "position": {"department": ..., "position": ...}.

Описанный этап является последним в алгоритме обработки запроса пользователя. После него сфор- мированный структурированный запрос выполня- ется СУБД, и информация из профилей сотрудни- ков, удовлетворяющих запросу, возвращается пользователю через графический интерфейс.

Программная реализация системы

Для реализации системы поиска был выбран язык программирования Python. Это высокоуров- невый язык программирования общего назначе- ния, ориентированный на повышение производи- тельности разработчика и читаемости кода. Син- таксис языка минималистичен, но его стандартная библиотека включает большой объем полезных

функций. Язык выделяется наличием мощных встроенных типов данных, таких как строки, мас- сивы и словари, что делает его удобным средством для разработки лингвистических процессоров.

Так как у авторов нет доступа к БД сотрудников ВШЭ, было решено составить достаточно точную ее локальную копию по информации, которую пре- доставляет веб-сайт НИУ ВШЭ. Для разбора раз- метки HTML использовался модуль lxml для языка Python. Он представляет собой интерфейс к биб- лиотеке lxml, предназначенной для построения моделей документов (document-object model, DOM) по HTML-разметке и манипуляции ими.

В качестве СУБД используется нереляционное хранилище MongoDB. Все данные в нем представ- лены в нотации JSON, а именно, в виде пар "ключ—значение" (key—value pairs) и упорядочен- ных множеств (collections). Так как все транзакции в ней подчиняются критериям ACID (Atomicity, Consistency, Isolation, Durability), работу подсистемы сбора информации можно в любой момент прервать и возобновить, при этом целостность данных не пострадает.

Для морфологического анализа слов использо- вана библиотека pymorphy2 — морфологический анализатор для русского языка. Для слов, включен- ных в словарь составных терминов, она показала точность распознавания более чем 95 %.

Для отображения пользовательского интерфейса использована библиотека PySide — программная среда для создания графических интерфейсов, ос- нованная на библиотеке Qt. При работе система демонстрирует пользователю СП его запроса и най- денные профили сотрудников, а на другой вкладке — более подробную информацию о выбранном со- труднике. Разработан также аналогичный веб-интер- фейс с использованием сетевого фреймворка Flask.

Система протестирована на работоспособность на компьютере с ОС Windows 8.1 и Ubuntu 14.04, процессором Intel Core i7 (4 ядра, тактовая частота 1.9 ГГц) и ОЗУ объемом 10 гигабайтов. Система также способна работать на системах Mac OS/Unix с необходимым окружением.

В качестве тестовых данных была использована информация о 8039 сотрудниках НИУ ВШЭ, сня- тая с их профилей на официальном сайте ([http:// www.hse.ru](http://www.hse.ru)) 1 ноября 2014 г. Она была предвари- тельно извлечена из Интернета и загружена в локаль- ную базу данных (MongoDB).

Пример результатов тестирования. Согласно определенному формату запросов составим мно- жество примеров запросов к системе: "сотрудник кафедры маркетинга фирмы"; "профессор, ведущий курс лекций по базам данных"; "стажер-исследо- ватель, закончивший до 2010 года специалитет Выс- шей школы экономики по специальности "Ме- неджмент". Прототип системы построил правиль- ные СП для данных запросов и успешно провел по ним поиск. По данным запросам система вывела 29, 12 и 1 результат соответственно.

Заключение

В ходе проведенного исследования разработан метод семантического поиска специалистов по ЕЯ-запросам, работающий с корпоративным хранилищем данных. На основе использования методологии теории К-представлений описан формат запросов, позволяющий определять набор критериев отбора специалистов и эффективно осуществлять по ним поиск. Предложена модифицированная модель лингвистической БД (ЛБД) по сравнению с моделями, введенными в работах [17, 21]. Эти два результата стали отправной точкой разработки и программной реализации на языке Python алгоритма построения СП-запросов пользователя, а из них — запросов к структурированному хранилищу информации.

Научная новизна исследования заключается в следующем:

- предложен новый практически полезный алгоритм разбора естественно-языковых выражений, рассматривающий группы слов для связывания их с информационными единицами и успешно обрабатывающий случаи некорректного пользовательского ввода;
- разработана расширенная модель ЛБД, упрощающая структуру словарей семантико-синтаксических фреймов;
- в язык семантических представлений, являющийся СК-языком в рассматриваемом концептуальном базисе, введены дополнительные выразительные механизмы.

Кроме того, следует заметить, что разработанный метод может быть применен для анализа текстов как на других естественных языках (например, на английском или немецком), так и для других профессиональных задач. Например, его перспективной областью применения может стать создание естественно-языковых интерфейсов для извлечения информации из онтологий посредством создания запросов на языке SPARQL.

В связи с этим наметим следующие направления для продолжения исследования:

- сбор данных о ЕЯ-запросах пользователей и изучение особенностей их обработки системой;
- поддержка более сложных на семантическом уровне форматов запросов;
- интернационализация и повышение универсальности алгоритма анализа ЕЯ для обеспечения возможности его применения в других задачах.

Проведенное исследование показало удобство использования на практике аппарата СК-языков, предложенного теорией К-представлений, и математической модели лингвистической базы данных.

Список литературы

1. Lin C.-Y., Cao N., Liu S. X., Papadimitriou S., Sun J., Yan X. SmallBlue Social Network Analysis for Expertise Search and Collective Intelligence // IEEE 25th International Conference on Data Engineering. ICDE '09. 2009. P. 1483—1486.
2. Farrell R. G., Pan F. Computing Similarities between Natural Language Descriptions of Knowledge and Skills // RC24060 Computer Science, IBM Research Division. 2006. V. 25.

3. Bull S., Greer J., McCalla G., Kettel L., Bowes J. User modeling in I-Help: What, why, when and how // User Modeling 2001. Springer, 2001. P. 117—126.
4. Staab S. Human language technologies for knowledge management // Intelligent Systems, IEEE. 2001. V. 16, N. 6. P. 84—94.
5. Maybury M. T. Expert Finding Systems. MITRE Technical Report. Bedford, Massachusetts: The MITRE Corporation, 2006. 52 p.
6. Kaufmann E., Bernstein A. How useful are natural language interfaces to the semantic web for casual end-users? // The Semantic Web. Springer. 2007. P. 281—294.
7. Chai J., Horvath V., Nicolov N., Stys M., Kambhatla N., Zadrozny W., Melville P. Natural Language Assistant — A Dialog System for Online Product Recommendation // AI Magazine. 2002. V. 23, N. 2. P. 63—76.
8. Berger H., Dittenbach M., Merkl D. An Adaptive Multilingual Interface for Tourism Information // International Journal of Electronic Business. 2004. V. 2, N. 5. P. 531—541.
9. Wang C., Xiong M., Zhou Q., Yu Y. PANTO: a portable NL-interface to ontologies // 4th European Semantic Web Conference Proceedings. Springer. 2007. P. 473—487.
10. Cimiano P., Haase P., Heizmann J., Mantel M., Studer P. Towards Portable Natural Language Interfaces to Knowledge Bases — the Case of the ORAKEL System // Data and Knowledge Engineering (DKE). 2008. V. 65, N. 2. P. 325—354.
11. Schwitler R. Creating and querying formal ontologies via controlled natural language // Applied Artificial Intelligence. 2010. V. 24, N. 1. P. 149—174.
12. Prince V., Roche M. Information Retrieval in Biomedicine: Natural Language Processing for Knowledge Integration. Hershey, Pennsylvania: IGI Global, 2009.
13. Захлебни И. В. Использование семантического анализа текстов для поиска специалистов // Supplementary Proceedings of AIST 2014. — CEUR Workshop Proceedings. 2014. V. 1197. P. 187—191 (in Russian).
14. Fomichov V. A. A Mathematical Model for Describing Structured Items of Conceptual Level // Informatica. An International Journal of Computing and Informatics (Slovenia). 1996. V. 20, N. 1. P. 5—32.
15. Фомичев В. А. Математические основы представления смысла текстов для разработки лингвистических информационных технологий. Часть I // Информационные технологии. 2002. № 10. С. 16—25.
16. Фомичев В. А. Математические основы представления смысла текстов для разработки лингвистических информационных технологий. Часть II // Информационные технологии. 2002. № 11. С. 34—45.
17. Фомичев В. А. Формализация проектирования лингвистических процессоров. М.: МАКС Пресс, 2005. 368 с.
18. Фомичев В. А. Понятие текстообразующей системы как компонент нового формального аппарата для проектирования лингвистических процессоров // Информационные технологии. 2005. № 8. С. 22—27.
19. Фомичев В. А. Новый метод преобразования естественно-языковых текстов в семантические представления // Информационные технологии. 2005. № 10. С. 25—35.
20. Фомичев В. А. Математические основы представления содержания посланий компьютерных интеллектуальных агентов. М.: ГУ-ВШЭ, ТЕИС, 2007. 176 с.
21. Fomichov V. A. Semantics-Oriented Natural Language Processing: Mathematical Models and Algorithms. New York, Dordrecht, Heidelberg, London; Springer, 2010. 354 p.
22. Правиков А. А., Фомичев В. А. Разработка рекомендательной системы с естественно-языковым интерфейсом на основе математических моделей семантических объектов // Бизнес-информатика. Междисциплинарный научно-практический журнал ГУ-ВШЭ. 2010. № 4(14). С. 3—11.
23. Fomichov V. A., Kirillov A. V. A Formal Model for Constructing Semantic Expansions of the Search Requests about the Achievements and Failures // Artificial Intelligence: Methodology, Systems, and Applications / Ed. by A. Ramsay, G. Agre. Lecture Notes in Computer Science. V. 7557. Berlin, Heidelberg: Springer, 2012. P. 296—304.
24. Razorenov A. A., Fomichov V. A. The Design of a Natural Language Interface for File System Operations on the Basis of a Structured Meanings Model // Procedia Computer Science, Elsevier. 2014. V. 31. P. 1005—1011; open access, URL: <http://authors.elsevier.com/sd/article/S1877050914005304>.
25. Fomichov V. A. Theory of K-representations as a Comprehensive Formal Framework for Developing a Multilingual Semantic Web // Informatica. An Intern. Journal of Computing and Informatics (Slovenia). 2010. V. 34, N. 3. P. 387—396.

Development of a Method for Semantic Search of Specialists in Corporate Databases Using Natural Language Queries

In the context of the studies on natural language processing, this paper substantiates the topicality of semantic search for specialists based on natural language queries. It also states the requirements to applied computer systems aimed at solving this problem, such as the necessity to extensively use semantics of natural language and structure of an available corporate database in the process of search. The principal ideas for development of such computer systems are set forth. A significant distinguished feature of the proposed method is the usage of SK-languages (standard knowledge languages), introduced in the V. A. Fomichov's theory of K-representations (knowledge representations), for formalizing semantics of natural language queries and for reflecting semantics of words and short word combinations in linguistic databases. The described method underpinned the design of a semantic search system ExpSearch, it was implemented in the programming language Python.

Keywords: semantic search of specialists, natural language query, natural language query semantic representation, theory of K-representations, SK-language, conceptual basis, linguistic database, corporate database, semantic-syntactic analyzer, Python

References

1. Lin C.-Y., Cao N. et al. SmallBlue: Social Network Analysis for Expertise Search and Collective Intelligence. *IEEE 25th International Conference on Data Engineering. ICDE'09. IEEE*. 2009. P. 1483—1486.
2. Farrell R. G., Pan F. Computing Similarities between Natural Language Descriptions of Knowledge and Skills. *RC24060 Computer Science, IBM Research Division*. 2006. V. 25.
3. Bull S., Greer J., McCalla G., Kettel L., Bowes J. User modelling in I-Help: What, why, when and how. *User Modeling* 2001. Springer, 2001. P. 117—126.
4. Staab S. Human language technologies for knowledge management. *Intelligent Systems, IEEE*. 2001. V. 16, N. 6. P. 84—94.
5. Maybury M. T. *Expert Finding Systems. MITRE Technical Report*. Bedford, Massachusetts: The MITRE Corporation, 2006. 52 p.
6. Kaufmann E., Bernstein A. How useful are natural language interfaces to the semantic web for casual end-users? *The Semantic Web*. Springer, 2007. P. 281—294.
7. Chai J., Horvath V., Nicolov N., Stys M., Kambhatla N., Zadrozny W., Melville P. Natural Language Assistant — A Dialog System for Online Product Recommendation. *AI Magazine*. 2002. V. 23. N. 2. P. 63—76.
8. Berger H., Dittenbach M., Merkl D. An Adaptive Multilingual Interface for Tourism Information. *International Journal of Electronic Business*. 2004. V. 2, N. 5. P. 531—541.
9. Wang C., Xiong M., Zhou Q., Yu Y. PANTO: a portable NL-interface to ontologies. *4th European Semantic Web Conference Proceedings*. Springer. 2007. P. 473—487.
10. Cimiano P., Haase P., Heizmann J., Mantel M., Studer P. Towards Portable Natural Language Interfaces to Knowledge Bases — the Case of the ORAKEL System. *Data and Knowledge Engineering (DKE)*. 2008. V. 65. N. 2. P. 325—354.
11. Schwitter R. Creating and querying formal ontologies via controlled natural language. *Applied Artificial Intelligence*. 2010. V. 24, N. 1. P. 149—174.
12. Prince V., Roche M. *Information Retrieval in Biomedicine: Natural Language Processing for Knowledge Integration*. Hershey, Pennsylvania: IGI Global, 2009.
13. Zakhlebin I. V. Ispolzovanie semanticheskogo analiza tekstov dlya poiska spetsialistov. *Supplementary Proceedings of AIST 2014*. — CEUR Workshop Proceedings. 2014. V. 1197. P. 187—191 (in Russian).
14. Fomichov V. A. A Mathematical Model for Describing Structured Items of Conceptual Level. *Informatica. An International Journal of Computing and Informatics (Slovenia)*. 1996. V. 20. N. 1. P. 5—32.
15. Fomichov V. A. Matematicheskie osnovy predstavleniya smysla tekstov dlya razrabotki lingvisticheskikh informatsionnykh tekhnologii. Part I. *Informatsionnye Tekhnologii*. 2002. N. 10. P. 16—25 (in Russian).
16. Fomichov V. A. Matematicheskie osnovy predstavleniya smysla tekstov dlya razrabotki lingvisticheskikh informatsionnykh tekhnologii. Part II. *Informatsionnye Tekhnologii*. 2002. N. 11. P. 34—45 (in Russian).
17. Fomichov V. A. *Formalizatsiya proektirovaniya lingvisticheskikh protsessorov*. M.: MAKS Press, 2005. 368 p.
18. Fomichov V. A. Ponyatie tekstoobrazuyshchei sistemy kak component formalnogo apparata dlya proektirovaniya lingvisticheskikh protsessorov. *Informatsionnye Tekhnologii* 2005. N. 8. P. 22—27 (in Russian).
19. Fomichov V. A. Novyi metod preobrazovaniya yestestvenno-yazykovykh tekstov v semanticheskie predstavleniya. *Informatsionnye Tekhnologii*. 2005. N. 10. P. 25—35 (in Russian).
20. Fomichov V. A. *Matematicheskie osnovy predstavleniya soderezhaniya poslanii kompyuternykh intellektualnykh agentov*. Moscow: State University — Higher School of Economics, Publishing House "TEIS". 2007. 176 p.
21. Fomichov V. A. *Semantics-Oriented Natural Language Processing: Mathematical Models and Algorithms*. New York, Dordrecht, Heidelberg, London; Springer, 2010. 354 p.
22. Pravikov A. A., Fomichov V. A. Razrabotka rekomendatelnoi sistemy s yestestvenno-yazykovym interfeisom na osnove matematicheskikh modelei semanticheskikh ob'ektov. *Biznes-informatika. Mezhdistsiplinarnyi nauchno-prakticheskii zhurnal*. 2010. N. 4(14). P. 3—11.
23. Fomichov V. A., Kirillov A. V. *A Formal Model for Constructing Semantic Expansions of the Search Requests about the Achievements and Failures*. Artificial Intelligence: Methodology, Systems, and Applications. Ed. by A. Ramsay, G. Agre. Lecture Notes in Computer Science. V. 7557. Berlin, Heidelberg: Springer, 2012. P. 296—304.
24. Razorenov A. A., Fomichov V. A. The Design of a Natural Language Interface for File System Operations on the Basis of a Structured Meanings Model. *Procedia Computer Science*, Elsevier. 2014. V. 31. P. 1005—1011; open access, URL: <http://authors.elsevier.com/sd/article/S1877050914005304>.
25. Fomichov V. A. Theory of K-representations as a Comprehensive Formal Framework for Developing a Multilingual Semantic Web. *Informatica. An Intern. Journal of Computing and Informatics (Slovenia)*. 2010. V. 34, N. 3. P. 387—396.

В. П. Кулагин, д-р техн. наук, проф., e-mail: vkulagin@hse.ru,
 Национальный исследовательский университет "Высшая школа экономики",
В. Я. Цветков, д-р техн. наук, проф., советник ректората, e-mail: cvj2@mail.ru,
 Московский государственный технический университет радиотехники и автоматики,
М. П. Лапчинская, канд. техн. наук, доц., e-mail: la.margo@miigaik.ru,
 Московский государственный университет геодезии и картографии

Идентификация морских объектов при сильном волнении водной поверхности

Идентификация является одной из главных задач, решаемых дистанционными методами и средствами, установленными на авиационных и космических платформах. Рассмотрен метод описания и анализа морской поверхности, характеризующейся сильным волнением. При этом предусматривается применение автоматизированного распознавания изображений с использованием характерных признаков, таких как размеры производных элементов текстуры, которые характеризуются нечеткостью линейного размера. В основу метода положен метод нечетких множеств. Метод позволяет получить модели морской поверхности с сильным волнением, которые дают возможность идентифицировать морские объекты по цифровым изображениям. На изображение накладывается модель поверхности, которая вычитается из исходного изображения. На оставшемся "очищенном" изображении идентификация морского объекта осуществляется как на поверхности без волнения.

Ключевые слова: идентификация, моделирование, морская поверхность с волнением, изображения, нечеткие множества

Введение

Идентификация в рамках данного проекта включает обнаружение и распознавание морских объектов, движущихся над морской поверхностью с волнением, а также измерение их параметров. Идентификация широко применяется при мониторинге морских судов [1] и является одной из главных задач, решаемых дистанционными методами и средствами, установленными на авиационных и космических платформах. Существуют разные подходы к решению этой задачи. В работе [2] применен подход, основанный на построении модели окружающей обстановки объекта, расположенного на морской поверхности. Для анализа информационных процессов как на уровне динамических (надводных и воздушных) объектов в зоне наблюдения, так и на уровне объектов, структурирующих входные информационные потоки о протекающих в модели процессах, используются мобильные программные агенты. В работах [3–5] описываются подходы идентификации морского объекта с учетом его скорости, использующие нелинейную модель, построенную на основе радиолокационных снимков. В работах [6, 7] рассматриваются алгоритмы оценки параметров динамической модели морского объекта с помощью расширенного фильтра Калмана. Решение задач распознавания морских объектов приводит к необходимости разработки новых методов синтеза и обработки зондирующих сигналов, при которых для протяженных объектов угловые и амплитудные шумы могут служить дополнительными информационными источниками, используемыми

для классификации и распознавания. При этом в работах [8, 9] отмечается необходимость применения методов, использующих данные, полученные со спутников, для управления морскими объектами и их идентификацией.

В итоге, с высокой степенью уверенности можно утверждать, что концептуально указанные задачи решаются построением информационной модели самого объекта и модели некоторого информационного окружения объекта [10, 11]. Однако общим недостатком указанных подходов является исключение фактора волнения морской поверхности при идентификации. В данной статье решается именно этот вопрос.

Способ представления морской поверхности

Изображения морской поверхности представляют собой текстурно-однородные поля вида $F(\Delta T, \Delta L_x, \Delta L_y)$, где ΔT — переменная, характеризующая оптическую плотность производных элементов; $(\Delta L_x, \Delta L_y)$ — переменные, описывающие пространственное расположение производных элементов текстурно-однородного поля для определенного подкласса подстилающей поверхности.

Данная совокупность переменных описывает геометрические, статические и физические свойства производных элементов текстуры и представляет собой интегральный признак текстуры изображения.

Эксперименты по распознаванию текстурно-однородных объектов на примере фотоизображений морской поверхности показали, что основной особенностью интегрального признака текстуры

фотоизображения являются *физическая* и *лингвистическая неопределенности* прямых визуальных дешифровочных признаков, проявляющиеся в виде различных нечетких свойств, в частности, при количественной интерпретации дешифровщиком линейных и площадных размеров производных элементов возможных ситуаций на изображении морской поверхности с волнением для таких подклассов, как "Рябь", "Рябь-на-ряби", "Слик", "Пена". Эти фотоизображения соответствуют определенному подклассу класса "Морская поверхность" с ярко выраженными структурными особенностями. Изучение этих изображений позволило выделить нечеткие свойства дешифровочного признака текстуры изображения "Цифровой муар" [12].

Рассмотрим информационную ситуацию [10], которую условно назовем "Волнение" и которая характеризуется набором таких параметров, как длина волны волнового холма λ , длина гребня волнового холма T , площадь волнового холма S , ориентация волнового холма α .

Указанные параметры при одних и тех же метеорологических условиях не имеют единственного значения, а характеризуются определенным разбросом. Параметры, имеющие неопределенность при измерениях, называются *нечеткими параметрами*. Данные параметры будем обозначать в дальнейшем волнистой чертой сверху: $\tilde{\lambda}$, \tilde{T} , \tilde{S} , $\tilde{\alpha}$.

Неопределенность приводит к неоднозначности компьютерного распознавания физических объектов и явлений на подстилающей поверхности при использовании различных групп текстурных признаков. При фотосъемке реальным элементам на морской поверхности будут соответствовать элементы снимка, которые в дальнейшем назовем *элементами текстуры*. Эти элементы делятся на производные, или базисные, и производные, которые формируются на основе базисных. Производные элементы играют роль информационных единиц, из которых формируют модели. На рис. 1 показан фрагмент реальной ситуации на морской поверхности и выделены наиболее важные характеристики ситуации "Волнение".

При применении автоматизированного распознавания изображений с использованием характерных признаков, таких как размеры производных элементов текстуры, исходные нечеткие параметры длины волны и длины гребня волнового холма обуславливают нечеткость линейного размера производного элемента текстуры p .

Подход к идентификации объекта с использованием муаровых картин

Волнение на море, показанное фрагментарно на рис. 1, относится к муаровым картинкам. Муаровой картиной называют картину, получаемую при взаимодействии двух групп периодических структур в изображении. При формировании муаровых картин морской поверхности угол поворота α является не-

четким. Информационной ситуации "Волнение" на море будет соответствовать информационная ситуация "Цифровой муар" на снимке.

Различный контраст изображения также влияет на формирование набора параметров для ситуации "Цифровой муар", рассчитываемого распознающей программой. При формировании муаровых картин на снимке участвуют не только производные элементы текстуры, соответствующие физическим объектам, но и еще два дополнительных типа производных элементов текстуры, обусловленные спецификой формирования изображения.

Первый дополнительный тип производных элементов — *фон*, который соответствует физическим элементарным объектам. Например, если участки изображения волновых холмов отобразить белым цветом (рис. 2, а), то фоновые производные элементы будут соответствовать негативному отображению волновых холмов (рис. 2, б).

Второй дополнительный тип производных элементов — *зернистость* изображения. Например, если отобразить производные элементы текстуры и, соответственно, производные элементы фона тонкой линией, то производные элементы зернистости накладываются на указанные производные элементы (схематично они представлены на рис. 2, в).

Таким образом, в формировании муаровых картин участвуют три типа производных элементов,

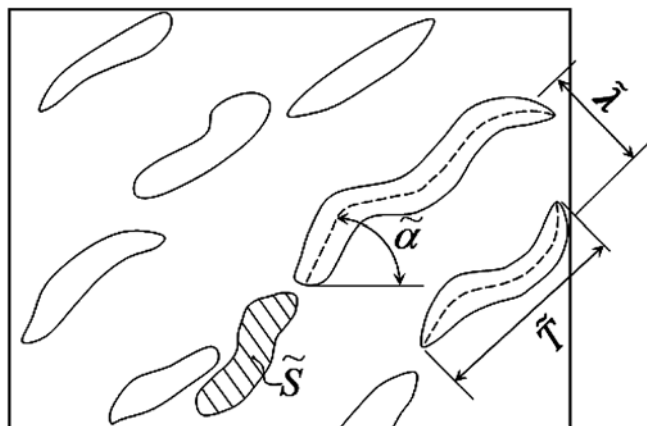


Рис. 1. Нечеткие параметры производного элемента текстуры изображения морской поверхности

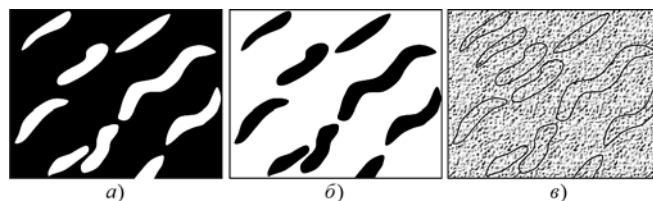


Рис. 2. Типы производных элементов текстуры p изображения, участвующих в формировании муаровых картин морской поверхности

из которых строятся производные (зависимые) элементы:

- первый тип — производные элементы текстуры изображения \tilde{p}_i , соответствующие структурным элементам в виде элементарных физических объектов класса/подкласса подстилающей поверхности Земли (рис. 2, а);
- второй тип — производные элементы фона \tilde{p}_f от производных элементов текстуры \tilde{p}_i изображения, т. е. $\tilde{p}_f = f(\tilde{p}_i)$ (рис. 2, б). Элементы фона занимают основной объем изображения и создают существенные помехи для идентификации объектов, сопоставимых с размером анализируемого участка;
- третий тип — производные элементы зернистости изображения \tilde{p}_z (рис. 2, в), которые являются самыми мелкими на изображении и создают помехи для идентификации малых морских объектов.

Теория нечетких множеств [13] позволяет описать лингвистические свойства любых объектов, которым нельзя однозначно дать количественную оценку. При использовании нечетких множеств нужно учитывать все факторы, формирующие изображение. В соответствии с вышеприведенными

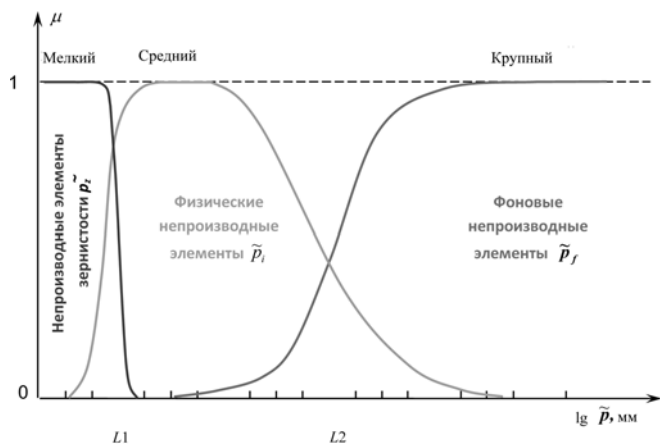


Рис. 3. Функции принадлежности непроизводного элемента p_i , p_f , p_z

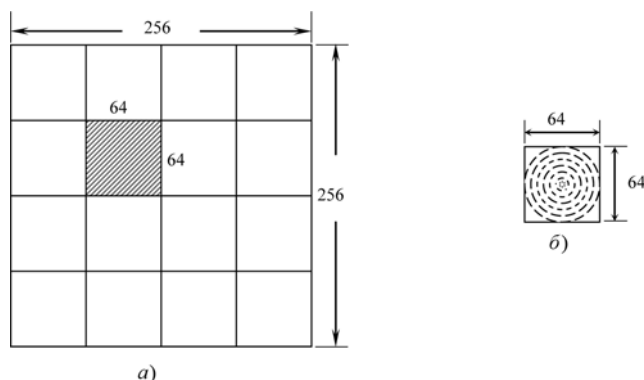


Рис. 4. Размеры анализируемых участков: цифровое изображение (а) и изображение цифрового муара (б) в пикселях

рассуждениями для информационной ситуации "Цифровой муар" необходимо учитывать три типа производных элементов, определенных выше. Согласно теории нечетких множеств данные типы производных элементов (\tilde{p}_i , \tilde{p}_f , \tilde{p}_z) можно задать тремя функциями принадлежности (рис. 3). Здесь: μ — значение функции принадлежности, которая, как и вероятность, лежит в пределах от 0 до 1 и описывает множество вероятностей для разных аргументов нечеткой величины. Выделенные на рис. 3 области, отмеченные как "мелкий", "средний" и "крупный", соответствуют диапазонам изменения описанных выше производных элементов — зернистость, объект (волна) и фон соответственно.

Границам областей "мелкий", "средний" и "крупный" соответствуют параметры L_1 и L_2 , которые имеют существенный физический смысл. Во-первых, определяя границы между зернистостью, фоном и изображением объекта, данные параметры задают участок наиболее вероятной идентификации объекта. Кроме этого, параметры L_1 и L_2 задают размер волн, что дает возможность ввести понятие и критерий "Большого волнения". Под большим волнением будем понимать такую характеристическую величину ($L_2 - L_1$), значение которой соизмеримо по масштабу с поперечными размерами идентифицируемого объекта.

Точность автоматизированного расчета вектора параметров текстурного признака "Цифровой муар" зависит от качества формируемой муаровой картины на фрагменте цифрового изображения, которое, в свою очередь, зависит от площади производного элемента $W = (\Delta L_x \times \Delta L_y)$.

Согласно предложенному в работе [14] алгоритму распознавания цифровых черно-белых изображений морской поверхности площадь анализируемых участков может быть равна 256×256 пикселей и больше (рис. 4, а), а площадь обрабатываемого в данный момент цифрового изображения может составлять не более 64×64 пикселей (рис. 4, б). Данный прием обработки цифрового изображения аналогичен известному методу "скользящего среднего" — в каждый момент времени обрабатывается только один участок изображения, но положение обрабатываемого участка на цифровом изображении меняется по определенному закону. Такой подход существенно (на порядок) сокращает объем обрабатываемой в конкретное время информации, скорость обработки данных повышается, а алгоритм обработки делается независимым от площади анализируемого изображения.

При анализе изображения следует учитывать, что масштаб снимка определяет число текстурных элементов и размер самого текстурного элемента, приведенных к единице площади. Экспериментальные исследования показали [12] (см. таблицу), что можно сформировать текстурные признаки объекта, имея в распоряжении лишь фрагмент

Экспериментальные оценки распознавания

Число производных элементов на участке цифрового изображения	Качество распознавания
656 389 106 32 23	Муаровая картина сформирована идеально
3 1	Муаровая картина сформирована некачественно
0,87 (часть) 0,48 (часть)	Муаровая картина сформирована плохо
0,26 (часть)	Муаровая картина не сформирована

цифрового изображения текстурного признака производного элемента.

Таким образом, качество автоматизированного анализа и распознавания изображений зависит от числа производных элементов на анализируемом фрагменте цифрового изображения.

Учет особенностей нечеткой текстуры изображения дает возможность применять статистические методы для формирования набора признаков при распознавании. На основе изложенного предлагается класс "Морская поверхность" описать тремя видами функций принадлежности, которые характеризуют зернистость изображения, фон и объект идентификации.

Эксперименты над муаровыми изображениями для текстур различных подклассов изображения морской поверхности показали, что из трех типов производных элементов, участвующих в формировании муаровых картин, наибольшей визуализацией обладает тип производных элементов текстуры изображения, поскольку обладает наибольшим контрастом составляющих его элементов. Однако вклад производных элементов фона \tilde{p}_f и производных элементов зернистости \tilde{p}_z изображения в формирование муаровой картины также существует, и он существенно различается для разных подклассов изображений подстилающей поверхности, что показали фотоизображения морской поверхности.

Заключение

Достоинство описанного подхода заключается в возможности математической обработки текстурных признаков, для которых невозможно дать единственную количественную оценку. Это позволяет перейти от качественного анализа морской поверхности с волнением к количественно-качественному анализу. Данный подход является составной частью координатного обеспечения решения задач мониторинга [15], что в дальнейшем может стать основой для автоматизированного распознавания и иденти-

фикации морских объектов на поверхности с волнением. Такой подход позволяет получать модели поверхности с волнением, которые дают возможность идентифицировать морские объекты по имеющимся изображениям, что позволяет создавать каталоги волнения для волн разной балльности. Другими словами, из имеющегося изображения "вычитается" модель морской поверхности с волнением, что приводит к "очищенному" изображению и более точной идентификации морского объекта.

Дальнейшим развитием данного метода является разработка автоматизированных методов распознавания морских объектов на основе каталогов подстилающей поверхности, в том числе с использованием тайловой организации визуальной информации [16].

Список литературы

1. Бармин И. В., Савиных В. П., Цветков В. Я., Затыгало В. В. Мониторинг загрязнений моря судами по данным дистанционного зондирования // Морской сборник. 2013. Т. 1998, № 9. С. 41—49.
2. Павлыгин Э. Д., Соснин П. И. Многоагентное моделирование и визуализация окружающей обстановки морского судна // Автоматизация процессов управления. 2010. № 2. С. 3—12.
3. Skjetne R., Smogeli O. N., Fossen T. I. A Nonlinear Ship Manoeuvring Model: Identification and adaptive control with experiments for a model ship // Modeling, identification and control. 2004. V. 25, N. 1. P. 3—27.
4. Горобец В. Н., Гутник В. Г., Зотов С. М., Кивва Ф. В., Шапиро А. А. Математическая модель радиолокационного образа корабля на морском волнении // Радиофизика та електроніка. 2011. Т. 2(16). № 4. С. 60—65.
5. Skjetne R., Smogeli O. N., Fossen T. I. A Nonlinear Ship Manoeuvring Model: Identification and adaptive control with experiments for a model ship // Modeling, identification and control. 2004. V. 25, N. 1. P. 3—27.
6. Chang S. J. Vessel identification and monitoring systems for maritime security // Security Technology, 2003. Proceedings. IEEE 37th Annual 2003 International Carnahan Conference on. IEEE, 2003. P. 66—70.
7. Fossen T. I., Sagatun S. I., Sorensen A. J. Identification of dynamically positioned ships // Control Engineering Practice. 1996. V. 4, N. 3. P. 369—376.
8. Buckley G. W. et al. Control system for a marine vessel: пат. 6273771 США. 2001.
9. Clifford P. J., Hart N. R., Meulman C. B. Satellite system for vessel identification: пат. 7483672 США. 2009.
10. Tsvetkov V. Ya. Information Situation and Information Position as a Management Tool // European Researcher. 2012. V. 36, N. 12-1. P. 2166—2170.
11. Tsvetkov V. Ya. Semantic environment of information units // European Researcher. 2014. V. 76, N. 6-1. P. 1059—1065.
12. Лапчинская М. П. Алгоритм классификации однородных текстур земной поверхности, основанный на построении функций принадлежности // Международный научно-технический и производственный журнал "Науки о Земле". 2012. № 4. С. 42—53.
13. Кофман А. Введение в теорию нечетких множеств. М.: Радио и связь. 1982. 432 с.
14. Лапчинская М. П. Метод цифрового муара. // Геодезия и аэрофотосъёмка. 1997. № 5. С. 86—113.
15. Савиных В. П. Система получения координатно-временной информации для решения задач мониторинга // Международный научно-технический и производственный журнал "Науки о Земле". 2012. Выпуск 03. С. 5—10.
16. Цветков В. Я., Лобанов А. А., Матчин В. Т., Железняков В. А. Обновление банков данных пространственной информации // Информатизация образования и науки. 2015. № 1 (25). С. 128—136.

V. P. Kulagin¹, Prof., e-mail: vkulagin@hse.ru,
V. Ya. Tsvetkov², Prof., e-mail: svj2@mail.ru, M. P. Lapchinskaya³
¹Higher School of Economics (National Research University)

²Moscow State Technical University of Radio Engineering, Electronics and Automation

³Moscow State University of Geodesy and Cartography

Identification of Ships in Heavy Seas

Identification is one of the main tasks of remote methods and tools that are installed on aircraft and space platforms. The article describes a method of description and analysis of the sea surface, characterized by a strong emotion. This will involve the application of automated image recognition using characteristic features, such as the dimensions of the basic elements of texture, which are characterized by vagueness of linear size. The method is based on a method of fuzzy sets. The method allows to obtain a model of the sea surface with strong agitation, which allow us to identify marine objects in digital images. On the image surface model, which is subtracted from the original image. The remaining "clean" the image identification of the marine facility is not a rough surface.

Keywords: identification, modeling, rough sea surface, images, fuzzy sets

References

1. Barmin I. V., Saviny'kh V. P., Tsvetkov V. Ya., Zatiagalova V. V. Monitoring zagriaznenii' moria sudami po danny'm distantsionnogo zondirovaniia. *Morskoi' sbornik*. 2013. V. 1998, N. 9. P. 41–49.
2. Pavly'gin E'. D., Sosnin P. I. Mnogoagentnoe modelirovanie i vizualizatsiia okruzhaiushchei' obstanovki morskogo sudna. *Avtomatizatsiia protsessov upravleniia*. 2010. N. 2. P. 3–12.
3. Skjetne R., Smogeli O. N., Fossen T. I. A Nonlinear Ship Manoeuvring Model: Identification and adaptive control with experiments for a model ship. *Modeling, identification ad control*. 2004. V. 25, N. 1. P. 3–27.
4. Gorobets V. N., Gutnik V. G., Zotov S. M., Kivva F. V., Shapiro A. A. Matematicheskaia model' radiolokatsionnogo obraza korablia na morskomo volnenii. *Radiofizika ta elektronika*, 2011. V. 2 (16). N. 4. P. 60–65.
5. Skjetne R., Smogeli O. N., Fossen T. I. A Nonlinear Ship Manoeuvring Model: Identification and adaptive control with experiments for a model ship. *Modeling, identification ad control*. 2004. V. 25, N. 1. P. 3–27.
6. Chang S. J. Vessel identification and monitoring systems for maritime security. *Security Technology*, 2003. Proceedings. IEEE 37th Annual 2003 International Carnahan Conference on. IEEE, 2003. P. 66–70.
7. Fossen T. I., Sagatun S. I., Sorensen A. J. Identification of dynamically positioned ships. *Control Engineering Practice*. 1996. V. 4. N. 3. P. 369–376.
8. Buckley G. W. et al. Control system for a marine vessel: Pat. 6273771 USA. 2001.
9. Clifford P. J., Hart N. R., Meulman C. B. Satellite system for vessel identification: Pat.7483672 USA. 2009.
10. Tsvetkov V. Ya. Information Situation and Information Position as a Management Tool. *European Researcher*. 2012. V. 36, N. 12-1. P. 2166–2170.
11. Tsvetkov V. Ya. Semantic environment of information units. *European Researcher*. 2014. V. 76, N. 6-1. P. 1059–1065.
12. Lapchinskaya M. P. Algoritm klassifikatsii odnorodny'kh tekstur zemnoi' poverkhnosti, osnovanny'i na postroenii funktsii' pri nadlezhnosti. *Mezhdunarodny'i' nauchno-tekhnicheskii' i proizvodstvenny'i' zhurnal "NAUKI O ZEMLE"*. 2012. N. 4. P. 42–53.
13. Kofman A. *Vvedenie v teoriu nechetkikh mnozhestv*. M.: Radio i sviaz', 1982. 432 p.
14. Lapchinskaya M. P. Metod tsifrovogo muara. *Geodeziiz i ae'rofotos'emka*. 1997. N. 5. P. 86–113.
15. Saviny'kh V. P. Sistema polucheniiia koordinatno-vremennoi' informatcii dlia resheniia zadach monitoringa. *Mezhdunarodny'i' nauchno-tekhnicheskii' i proizvodstvenny'i' zhurnal "Nauki o Zemle"*. 2012. Iss. 03. P. 5–10.
16. Tsvetkov V. Ya., Lobanov A. A., Matchin V. T., Zhelezniakov V. A. Obnovlenie bankov danny'kh prostranstvennoi' informatcii. *Infomatizatsiia obrazovaniia i nauki*. 2015. N. 1(25). P. 128–136.

УДК 004.94

С. В. Лучкова, канд. техн. наук, мл. науч. сотр., Т. О. Перемитина, канд. техн. наук, науч. сотр.,
И. Г. Ященко, канд. геол.-мин. наук, зав. лаб.

Федеральное государственное бюджетное учреждение науки Институт химии нефти СО РАН, Томск

Использование нечеткого моделирования для повышения репрезентативности информации на примере анализа характеристик нефти

Рассмотрена задача восстановления пропущенных значений в выборке о свойствах сернистых нефтей. Описывается алгоритм восстановления данных на основе нечеткого моделирования. Приведены результаты экспериментальных исследований для повышения репрезентативности выборки на основе неполных данных.

Ключевые слова: восстановление данных, нечеткое моделирование, физико-химические свойства нефти

Введение

Для изучения природных объектов анализируются данные, которые имеют сложную структуру и представлены многомерными выборками с пропущенными значениями. Существующая необходимость в анализе неполных данных предполагает их предварительную обработку, такую как восполнение недостающих значений для увеличения репрезентативности выборки. Это необходимо, так как в самом простом решении для анализа таких данных (исключить все записи с пропусками) существуют проблемы с ухудшением достоверности выводов анализа вследствие сокращения размера выборки и, как следствие, проявлением статистических смещений. Однако для первоначального анализа и выявления особенностей данных используются именно полные значения (выборки данных без пропущенных значений), которые позволяют оценить предстоящую работу и выбрать подходящие методы восстановления.

Одним из продуктивных подходов к решению проблемы о пропущенных значениях, касательно нефти, является применение методов нечеткого моделирования, которые позволяют наиболее точно восстановить недостающие данные за счет работы с выборками малой размерности и в условиях статистической неопределенности [1].

Этапы восстановления данных

Процедура восстановления данных предполагает трудоемкий предварительный процесс, который состоит из нескольких этапов:

1. Предварительный анализ исходных данных для оценки их полноты.
2. Формирование полной выборки.
3. Построение модели исследуемых данных по полной выборке.
4. Проверка адекватности построенной модели.
5. Восстановление данных на основе построенной модели.

Рассмотрим этапы подробнее.

Этап "Предварительный анализ" включает в себя не только оценку полноты данных, но также помогает выявить особенности объекта исследования, например, проследить средние значения характеристик, разброс значений характеристик, какие-либо взаимосвязи.

Этап "Формирование полной выборки" заключается в исключении всех записей из выборки, в которых есть пропущенные значения, а также позволяет принять решение о "случайных выбросах" в значениях характеристик, которые не являются характерными для данных и могут оказать значительное влияние на статистические величины. Для этого этапа могут быть использованы различные статистические методы, которые помогут понять исследуемые данные, например, по полной таблице

можно выявить корреляционные связи и главные факторы с помощью факторного анализа [2].

Этап "Построение модели" выполняется с помощью нечеткого моделирования, которое включает в себя два основных подэтапа: идентификацию структуры (НС) с заданной структурой [3], с треугольной функцией принадлежности для построения системы и методом эволюционной стратегии (ЭС) для настройки параметров системы [4, 5].

Идентификация структуры — это определение основных характеристик нечеткой модели (число нечетких правил, число лингвистических термов для входных и выходных переменных). Настройка же параметров НС — это определение неизвестных параметров антецедентов (условная часть — "ЕСЛИ") и консеквентов (заключительной части — "ТО") нечетких правил путем оптимизации работы нечеткой модели по заданному критерию.

Правила нечеткой модели типа сингльтон имеют следующий вид:

$$\text{правило } i: \text{ЕСЛИ } x_1 = A_{1i} \text{ И } x_2 = A_{2i} \text{ И } \dots \text{ И } x_m = A_{mi} \\ \text{ТО } y = r_i$$

где A_{ji} — лингвистический терм, которым оценивается переменная x_j , а выход y оценивается действительным числом r_i .

Модель осуществляет отображение $F: \mathfrak{R}^m \rightarrow \mathfrak{R}$, заменяя оператор нечеткой конъюнкции произведением, а оператор агрегации нечетких правил сложением. Отображение F для модели типа сингльтон определяется формулой

$$F(\mathbf{x}) = \frac{\sum_{j=1}^m r_j \prod_{i=1}^n \mu_{A_{ij}}(x_i)}{\sum_{j=1}^m \prod_{i=1}^n \mu_{A_{ij}}(x_i)},$$

где $\mathbf{x} = [x_1, \dots, x_n]^T \in \mathfrak{R}^n$ — значение i -го входа; $\mu_{A_{ij}}(x_j)$ — функция принадлежности лингвистического терма A_{ij} ; r_j — значение консеквента в j -м правиле.

Нечеткая система может быть представлена как $y = f(\mathbf{x}, \boldsymbol{\theta})$, где $\boldsymbol{\theta} = \|\theta_1, \dots, \theta_N\|$ — вектор параметров; N — сумма термов по каждому исследуемому параметру; y — скалярный выход системы.

Задача параметрической идентификации — определить неизвестные параметры антецедентов и консеквентов нечетких правил путем оптимизации работы нечеткой системы по заданному критерию.

Параметрическая идентификация рассматривается как процесс оптимизации нечеткой модели, который сводится к нахождению таких параметров нечеткой системы, при которых ошибка вывода была бы минимальной. При этом оценивается качество нечеткого вывода по значениям ошибки вывода — разницы между значениями выходной пере-



Рис. 1. Алгоритм восстановления данных

менной из таблицы наблюдений $f(x)$ и значениями $F(x)$, полученными нечеткой системой. В работе для оценки результатов вывода НС используется средняя квадратичная ошибка (СКО):

$$E(\theta) = \sqrt{\frac{\sum_i (f(x_i) - F(x_i, \theta))^2}{N}}$$

На рис. 1 представлен алгоритм восстановления данных, основанный на описанной модели.

Далее необходимо проверить "насколько корректно подобрана модель". Так, на **этапе "Проверка адекватности"** полученная модель проверяется с помощью метода скользящего экзамена [6], когда в полную таблицу будут построчно вводиться искусственные пропуски, затем значения будут восстанавливаться согласно предложенному алгоритму восстановления и сравниваться с исходными значениями. Такая проверка позволит сделать выводы о полученной модели и решить, подойдет ли она для дальнейшей обработки данных. Если же модель подходит, наступает следующий **этап — "Восстановление данных"** в полной выборке по имеющейся адекватной модели.

Объект исследования и постановка эксперимента

В условиях роста объемов добычи углеводородного сырья во всем мире и при существенном истощении запасов качественной нефти становятся актуальными вопросы освоения трудноизвлекае-

мых запасов нефти и изучение физико-химических свойств нефти для возможности прогнозирования ее качества в новых районах нефтедобычи [7]. К трудноизвлекаемым относятся сернистые нефти, для которых, согласно классификации нефтей [8], содержание серы от 3 мас. % и выше. Так, в целях апробации указанных методов восстановления пропущенных значений на основе информации из базы данных Института химии нефти СО РАН [9] была сформирована выборка из 548 образцов сернистой нефти по восьми характеристикам. Описание характеристик представлено в табл. 1.

Первоначальный анализ показал, что из всего массива данных общим объемом 4384 значений (матрица 548×8) пропущено 1221 значение (27,85 % общего количества данных). В табл. 2 приведено число пропусков для каждой из восьми характеристик. Данный предварительный анализ информации позволяет оценить возможности восстановления, исключив записи, дающие при анализе сильные статистические смещения.

Общее соотношение пропусков менее 30 %, что позволяет использовать вышеописанный метод восстановления данных с помощью нечеткого моделирования и ожидать успешного результата. Анализ же неполноты отдельных характеристик, таких как температура пласта и пластовое давление показывает, что данные параметры следует тщательно моделировать, так как практически 65 % данных отсутствует, что может привести к восстановлению данных с большой ошибкой вывода системы.

Таблица 1

Характеристики объекта исследования

№	Характеристики сернистой нефти	Среднее значение
1	Плотность, г/см ³	0,9238
2	Вязкость при 20 °С, мм ² /с	590,3
3	Содержание серы, % мас.	3,87
4	Содержание парафинов, % мас.	3,912
5	Содержание смол, % мас.	23,18
6	Содержание асфальтенов, % мас.	7,42
7	Температура пласта, °С	33,28
8	Пластовое давление, МПа	14,79

Таблица 2

Количество пропусков в характеристиках объекта исследования

№	Характеристики нефти	Число пропусков/(%)
1	Плотность, г/см ³	9/(1,64 %)
2	Вязкость при 20 °С, мм ² /с	191/(34,85 %)
3	Содержание серы, % мас.	0/0
4	Содержание парафинов, % мас.	121/(22,08 %)
5	Содержание смол, % мас.	116/(21,16 %)
6	Содержание асфальтенов, % мас.	79/(14,41 %)
7	Температура пласта, °С	349/(63,68 %)
8	Пластовое давление, МПа	356/(64,96 %)
	Общее соотношение (от 4384)	1221/(27,85 %)

Оптимальные параметры нечеткой системы

Число термов на каждый параметр	Средняя квадратичная ошибка вывода нечеткой системы
(5 3 3 6 3 4 3 3)	0,043547
(7 5 3 8 4 4 3 3)	0,051112
(6 4 3 7 4 5 4 3)	0,037304

Примечание. Число термов представлено для характеристик нефти в порядке следования в табл. 1.

Далее следует сформировать "полную" выборку для моделирования. Ввиду числа пропусков, описанных выше, в сформированной "полной" таблице всего 75 записей без пропусков (из 548), на основе которых будет проведено моделирование системы и проверка адекватности подобранной системы.

Моделирование системы

Нечеткая система с заданной структурой [3] предполагает, что для каждого показателя нефти необходимо подобрать оптимальное разбиение функцией принадлежности, за которое отвечает показатель "терм" в структуре нечеткой системы. Подробно способ подбора параметров был описан в работе [10]. Так, для полной выборки из 75 образцов нефти проведено моделирование для поиска наиболее оптимальных параметров нечеткой системы и метода эволюционной стратегии. После ряда экспериментов были получены результаты, часть из которых представлена в табл. 3.

Как видно из табл. 3, из представленных вариантов лучше всего подходит третий, который характеризуется разбиением на термы от 3 до 7. При этом были использованы следующие параметры метода эволюционной стратегии: алгоритм стратегии (20 + 50), 1000 итераций, алгоритм скрещивания — многоточечный, алгоритм селекции — элитарный, вероятность мутации — 0,15.

Результаты проверки построенной модели с помощью метода скользящего экзамена следующие:

- средняя квадратичная ошибка вывода НС — 0,02506;
- точность модели — 96,78 %.

Получив представленный результат, было принято решение, что подобранная модель имеет оптимальные параметры, а следовательно, можно перейти к основной задаче — повышению репрезентативности выборки сернистой нефти, другими словами, с помощью полученной модели необходимо восстановить пропущенные значения в исходной выборке. Результаты средних показателей характеристик нефти до и после восстановления представлены на рис. 2.

Как видно из рис. 2, видимое различие между данными есть только в показателе "Вязкость", значение которого в исследуемой выборке сернистой нефти варьируется от 1,071 до 70077,93 мм²/с, что и объясняет такое расхождение в значениях до и после восстановления. Для более точного восстановления данного параметра необходимо дополнительное исследование и, возможно, привлечение новой информации. Остальные же показатели находятся в допустимом интервале 5 %, что позволяет сделать выводы об успешном восстановлении и возможности дальнейшего использования методов восстановления данных для статистического анализа свойств не только сернистых нефтей, но и других видов трудноизвлекаемых нефтей.

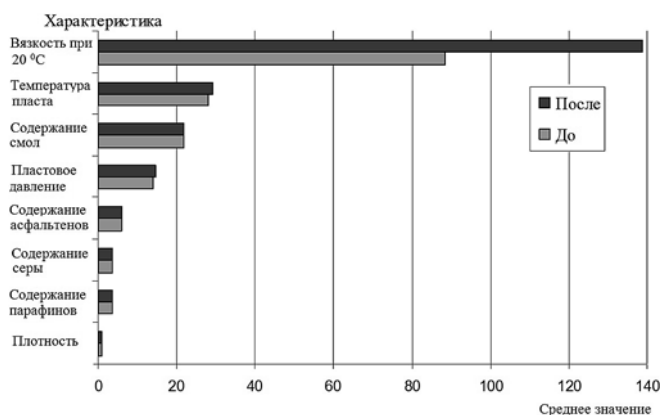


Рис. 2. Средние показатели характеристик сернистой нефти до и после восстановления

Заключение

В работе рассмотрены основные моменты моделирования и восстановления пропущенных данных на основе нечеткого моделирования с использованием информации об основных физико-химических свойствах сернистой нефти. Рассмотрены результаты моделирования объекта исследования в условиях малой размерности выборки. Проведена проверка полученной модели с помощью алгоритма скользящего экзамена, что позволило принять модель и повысить репрезентативность исходной выборки характеристик сернистой нефти. Восстановленная выборка позволит провести последующий анализ характеристик нефти, выявить новые закономерности изменения свойств нефти или подтвердить те, которые были получены в условиях статистической неопределенности.

Список литературы

1. Лучкова С. В., Перемитина Т. О. Применение программного комплекса "Нечеткая система на основе эволюционной стратегии" для задачи импутирования // Информационные технологии. 2013. № 2. С. 47—50.
2. Ходашинский И. А., Гнездилова В. Ю., Дудин П. А., Лавыгина А. В. Основанные на производных и метаэвристические методы идентификации параметров нечетких моделей // Труды VIII международной конференции "Идентификация систем и задачи управления" SICPRO '08 (Москва). 2009. С. 501—528.
3. Bahrami H., Abdechiri M., Meybodi M. R. Imperialist Competitive algorithm with adaptive colonies movement // I. J. Intelligent System and Applications. 2012. N. 2. P. 49—57.

4. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. М.: Горячая линия—Телеком, 2006. 383 с.

5. Hoche S., Wrobel S. A Comparative Evaluation of Feature Set Evolution Strategies for Multirelational Boosting // Proc. 13th Int. Conf. on ILP 2003.

6. Загоруйко Н. Г. Методы распознавания и их применение. М.: Сов. радио, 1972. 216 с.

7. Лисовский Н. Н., Халимов Э. М. О классификации трудноизвлекаемых запасов // Вестник ЦКР Роснедра. 2009. № 6. С. 33—35.

8. Ященко И. Г., Полищук Ю. М. Трудноизвлекаемые нефти: физико-химические свойства и закономерности размещения. Томск: В-Спектр. 2014. 154 с.

9. База данных "химия нефти и газа" [Электронный ресурс]. URL: <http://enrit.jpc.tsc.ru/basel.htm> (дата посещения 15.09.2014).

10. Лучкова С. В., Перемитина Т. О., Ященко И. Г. Программный комплекс восстановления пропущенных значений в многомерных данных на основе методов нечеткого моделирования // Программные продукты и системы. 2014. № 1. С. 80—86.

S. V. Luchkova, Junior Researcher, e-mail; sonetta27@gmail.com, T. O. Peremitina, Researcher, I. G. Yashchenko, Head of Laboratory, Institute of Petroleum Chemistry, Tomsk

Usage Fuzzy Modelling for Raising a Sample Representative on the Example the Analysis of Characteristics Oils

In this paper we consider data recovering process of multidimensional data for sulphurous oils sample with missing values. The algorithm of data recovery is described. Experimental researches results are presented. Results of experimental researches for increasing representative data values based on missing data are shown.

Keywords: fuzzy modeling, data recovered, physic-chemical properties of oils

References

1. Luchkova S. V., Peremitina T. O. Primenenie programmnoho kompleksa "Nechetkaja sistema na osnove jevoljucionnoj strategii" dlja zadachi imputirovanija. *Informacionnye tehnologii*. 2013. N. 2. P. 47—50.

2. Hodashinskij I. A., Gnezdilova V. Ju., Dudin P. A., Lavygina A. V. Osnovannye na proizvodnyh i metajevristicheskie metody identifikacii parametrov nechetkih modelej. *Trudy VIII mezhdunarodnoj konferencii "Identifikacija sistem i zadachi upravlenija" SICPRO '08* (Moscow). 2009. P. 501—528.

3. Bahrami H., Abdechiri M., Meybodi M. R. Imperialist Competitive algorithm with adaptive colonies movement. *I. J. Intelligent System and Applications*. 2012. N. 2. P. 49—57.

4. Rutkovskaja D., Piliń'skij M., Rutkovskij L. Nejrornyje seti, geneticheskie algoritmy i nechetkie sistemy. Moscow: Gorjachaja linija—Telekom, 2006. 383 p.

5. Hoche S., Wrobel S. A Comparative Evaluation of Feature Set Evolution Strategies for Multirelational Boosting. *Proc. 13th Int. Conf. on ILP 2003*.

6. Zagorujko N. G. Metody raspoznavanija i ih primenenie. Moscow: "Sov. radio", 1972. 216 p.

7. Lisovskij N. N., Halimov Je. M. O klassifikacii trudnoizvlekaemyh zapasov. *Vestnik CKR Rosnedra*. 2009. N. 6. P. 33—35.

8. Yashchenko I. G., Polishchuk Yu. M. Trudnoizvlekaemye nefiti: fiziko-himicheskie svojstva i zakonomernosti razmeshhenija. Tomsk: V-Spektr, 2014. 154 p.

9. Baza dannyh "himija nefiti i gaza". URL: <http://enrit.ipc.tsc.ru/base1.htm> (data poseshhenija 15.09.2014)

10. Luchkova S. V., Peremitina T. O., Yashchenko I. G. Programmnyj kompleks vosstanovlenija propushhennyh znachenij v mnogomernyh dannyh na osnove metodov nechetkogo modelirovanija. *Programmnye produkty i sistemy*. 2014. N. 1. P. 80—86.

МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ MODELING AND OPTIMIZATION

УДК 519.233

Б. Г. Кухаренко, канд. физ.-мат. наук, ст. науч. сотр., вед. науч. сотр.,
Институт машиноведения РАН, г. Москва, e-mail: kukharenko@imash.ru

М. О. Солнцева-Чалей, аспирант,
Московский физико-технический институт (ГУ), e-mail: solnceva.chalei@gmail.com

Применение моделей нелинейных динамических систем для анализа результатов кластеризации многомерных траекторий

Анализ кластеров траекторий, полученных методом полиномиальных регрессий, выполняется с помощью адаптивного нелинейного фильтра Калмана. Для решения задачи одновременной оценки неизвестных параметров и скрытых состояний нелинейной динамической системы применяется алгоритм ожидания и максимизации правдоподобия, использующий приближения гауссовых радиальных базисных функций для представления нелинейной динамики вектора состояний системы. Выделенные компоненты вектора состояний системы позволяют выявить тонкую структуру кластера в соответствии с мерой косинуса.

Ключевые слова: анализ данных, многомерные траектории, кластеризация, полиномиальная регрессия, адаптивный фильтр Калмана, сглаживатель Рауха, нелинейные динамические системы, радиальные базисные функции, алгоритм ожидания и максимизации правдоподобия

Введение

Задача кластеризации траекторий обусловлена необходимостью организации движения управляемых объектов. При рассмотрении траекторий движения в трехмерном пространстве наиболее естественно их разделение на кластеры согласно характерной форме траекторий. Такому подходу к выделению кластеров наилучшим образом отвечает метод полиномиальных регрессий, позволяющий оценить форму некоторой обобщенной траектории (маршрута) для каждого кластера [1]. Одновременные кластеризация и выравнивание траекторий движения самолетов по методу полиномиальных регрессий являются достаточно эффективной мерой для решения этой задачи [2].

Траектории объектов описываются векторами переменной длины, представляющими зависимость их координат от времени. Каждый вектор $\mathbf{s}_i \in \mathbb{R}^{N_i \times 1}$, $i = \overline{1, q}$ (одномерный временной ряд) состоит из последовательности измерений координатной зависимости $\mathbf{s}_i = \mathbf{s}_i(t)$ в моменты времени $\mathbf{t}_i \in \mathbb{R}^{N_i \times 1}$. Вектор \mathbf{s}_i моделируется регрессионной моделью

$$\mathbf{s}_i = \mathbf{T}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad (1)$$

где $\boldsymbol{\beta}$ — вектор коэффициентов регрессии размерности $(m + 1) \times 1$, и $\boldsymbol{\varepsilon}_i$ — гауссов шум с нулевым средним, а \mathbf{T}_i — регрессионная матрица. Матрица \mathbf{T}_i зависит от типа используемой регрессионной мо-

дели. В случае полиномиальной регрессии \mathbf{T}_i имеет вид стандартной матрицы Вандермонта

$$\mathbf{T}_i = [[1 \ t_i[n] \ (t_i[n])^2 \ \dots \ (t_i[n])^m], \ n = \overline{1, N_i}]. \quad (2)$$

Основой одновременных кластеризации и выравнивания является модель смеси регрессий, в которую вводятся четыре независимых параметра преобразований выравнивания и масштабирования во времени и пространстве $\{\Phi_l\} = \{a_l, b_l, c_l, d_l\}$ (параметры a_l и b_l описывают масштабирование и сдвиг во времени, а параметры c_l и d_l — масштабирование и смещение в пространстве измерений) [3]. Полиномиальная регрессия для одномерного случая имеет вид

$$\mathbf{s}_i = c_l \Upsilon_l \boldsymbol{\beta}_l + d_l + \boldsymbol{\varepsilon}_i, \quad (3)$$

где матрица Υ_l получается из матрицы \mathbf{T}_i (2) подстановкой $\mathbf{t}_i \rightarrow a_l \mathbf{t}_i - b_l$, параметры $\boldsymbol{\beta}_l$ определяют полиномиальную регрессию для траекторий из l -го кластера ($l = \overline{1, L}$); $\boldsymbol{\varepsilon}_i$ — гауссов шум с нулевым средним и дисперсией $\sigma_l^2 \mathbf{I}$. Поэтому распределение плотности условной вероятности имеет вид

$$p_l(\mathbf{s}_i | a_l, b_l, c_l, d_l) = \mathcal{N}(\mathbf{s}_i | c_l \Upsilon_l \boldsymbol{\beta}_l + d_l, \sigma_l^2 \mathbf{I}). \quad (4)$$

Плотность вероятности для кривой \mathbf{s}_i однозначно задается соответствующим множеством параметров $\{\Phi_l\}$, которые требуется определить. Задача кластеризации кривых решается как стандартная задача оценки значений скрытых переменных. Каждый из параметров преобразования в выражениях (3) и (4) рассматривается как характерная для \mathbf{s}_i случай-

ная переменная с заранее известным распределением вероятности для кластера. Параметры преобразования в (3) и (4) и параметры $\beta_l, l = \overline{1, L}$, полиномиальной регрессии оцениваются одновременно посредством алгоритма ожидания и максимизации правдоподобия (EM-алгоритма) [4].

Число кластеров L , определяемых методом полиномиальных регрессий, невелико и на практике не превышает 10. Поэтому для набора достаточно неоднородных траекторий, полученные кластеры также не являются однородными. Для неоднородного кластера полиномиальная регрессия является весьма общим представлением. Поэтому, чтобы выявить неоднородность кластеров траекторий самолетов, структуру кластеров анализируют с помощью моделей линейных динамических систем [3]. Для подтверждения выявленной тонкой структуры кластеров в настоящей работе анализ кластеров выполняется более общим методом моделей нелинейных динамических систем [5].

Определение тонкой структуры рассматриваемого кластера траекторий требует определения компонент, являющихся скелетными кривыми его подкластеров. Определение компонент скрытого состояния заданной размерности и вывод относительно линейности/нелинейности многомерных временных рядов могут быть сделаны в результате анализа этих временных рядов с помощью адаптивного нелинейного фильтра Калмана [6, 7]. При этом для удобства описания вместо матрицы траекторий (измерений) $\mathbf{Y} = [\mathbf{s}_1, \dots, \mathbf{s}_q]$, где $\mathbf{s}_1, \dots, \mathbf{s}_q$ — вектор-столбцы из (1) и (3), используется транспонированная матрица, которая имеет вид

$$[\mathbf{y}[1], \dots, \mathbf{y}[N]] = \mathbf{Y}^T \in \mathbb{R}^{q \times N}.$$

1. Адаптивный нелинейный фильтр Калмана

Для обучения нелинейного фильтра Калмана используется алгоритм ожидания и максимизации правдоподобия (*Expectation-Maximization algorithm* — EM-алгоритм) [5, 8]. Нелинейная динамическая система в дискретном времени описывает эволюцию состояния $\mathbf{x}[k+1] \leftarrow \mathbf{x}[k]$ на одном временном шаге и текущую связь состояния и входа $\mathbf{x}[k], \mathbf{u}[k]$ с наблюдениями $\mathbf{y}[k], k = \overline{1, N}$:

$$\left. \begin{aligned} \mathbf{x}[k+1] &= \mathbf{f}(\mathbf{x}[k], \mathbf{u}[k]) + \mathbf{v}[k], \\ \mathbf{y}[k] &= \mathbf{g}(\mathbf{x}[k], \mathbf{u}[k]) + \mathbf{w}[k], \end{aligned} \right\} \quad (5)$$

где $\mathbf{v}[k], k = \overline{1, N}$ и $\mathbf{w}[k], k = \overline{1, N}$ — гауссовы шумы с нулевым средним [6, 7]. Динамические системы непрерывного времени (в которых производные специфицируются как функции текущего состояния и входа) могут быть преобразованы в системы дискретного времени (5) посредством дискретизации (*sampling*) их состояния и входа [9]. В частности, для линейной системы непрерывного времени

$$\mathbf{x}(t) = \mathbf{A}_c \mathbf{x}(t) + \mathbf{B}_c \mathbf{u}(t)$$

при дискретизации с интервалом времени τ

$$\mathbf{x}_{k+1} = \mathbf{A} \mathbf{x}_k + \mathbf{B} \mathbf{u}_k,$$

динамическая матрица и матрица влияния имеют вид

$$\mathbf{A} = \sum_{k=0}^{\infty} \frac{\mathbf{A}_c^k \tau^k}{k!} = \exp(\mathbf{A}_c \tau)$$

и

$$\mathbf{B} = \mathbf{A}_c^{-1} (\mathbf{A} - \mathbf{I}) \mathbf{B}_c.$$

Вектор состояния $\mathbf{x}[k]$ эволюционирует в соответствии с нелинейной, но стационарной, марковской динамикой, производимой входом $\mathbf{u}[k], k = \overline{1, N}$ в присутствии шума $\mathbf{w}[k], k = \overline{1, N}$. Наблюдения $\mathbf{y}[k], k = \overline{1, N}$, нелинейные с шумом, но стационарные, и являются функцией текущего состояния $\mathbf{x}[k]$ и текущего входа $\mathbf{u}[k]$. Нелинейные вектор-функции $\mathbf{f}(\cdot)$ и $\mathbf{g}(\cdot)$ считаются дифференцируемыми. Ниже на E -шаге (EM-алгоритма) для оценки приблизительного распределения скрытых состояний нелинейной системы (2) используется расширенный сглаживатель Рауха, а на M -шаге для нелинейной регрессии вектор-функций $\mathbf{f}(\cdot)$ и $\mathbf{g}(\cdot)$ — разложение по радиальным базисным функциям (*radial basis function* — RBF) [5, 8].

Два условных распределения вероятности

$$\begin{aligned} P(\mathbf{x}[k] | \mathbf{u}[1], \dots, \mathbf{u}[N], \mathbf{y}[1], \dots, \mathbf{y}[N]), k = \overline{1, N}, \\ P(\mathbf{x}[k], \mathbf{x}[k+1] | \mathbf{u}[1], \dots, \mathbf{u}[N], \mathbf{y}[1], \dots, \mathbf{y}[N]), k = \overline{1, N-1}, \end{aligned} \quad (6)$$

используются на E -шаге для определения последовательности скрытых состояний нелинейной системы (5) на основе последовательностей наблюдений $\{\mathbf{y}[k], k = \overline{1, N}\}$ и входов $\{\mathbf{u}[k], k = \overline{1, N}\}$. Условные распределения (6) являются не гауссовыми, поэтому уравнения вывода не могут быть представлены в замкнутой форме. Более того, объемы вычислений растут экспоненциально с увеличением длины временных рядов. Расширенный сглаживатель Рауха аппроксимирует стационарную нелинейную динамическую систему (2) нестационарной линейной системой [9]. Он применяет стандартный сглаживатель Рауха к локально линеаризованной нелинейной системе. В каждой точке $\bar{\mathbf{x}}$ в пространстве состояний \mathbf{x} , производные вектор-функций $\mathbf{f}(\cdot)$ и $\mathbf{g}(\cdot)$ определяют матрицы

$$\mathbf{F}_{\bar{\mathbf{x}}} = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\mathbf{x}=\bar{\mathbf{x}}} \quad \text{и} \quad \mathbf{G}_{\bar{\mathbf{x}}} = \left. \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \right|_{\mathbf{x}=\bar{\mathbf{x}}},$$

соответственно. Уравнения (2) линеаризуются в окрестности $\bar{\mathbf{x}}[k]$ средней текущей отфильтрованной (а не сглаженной) оценки состояния $\mathbf{x}[k]$ (в момент времени k). Аналогично линеаризуется уравнение для наблюдения. Эти линеаризации дают уравнения

$$\left. \begin{aligned} \mathbf{x}[k+1] &\approx \mathbf{f}(\bar{\mathbf{x}}[k], \mathbf{u}[k]) + \mathbf{F}_{\bar{\mathbf{x}}[k]} (\mathbf{x}[k] - \bar{\mathbf{x}}[k]) + \mathbf{w}[k], \\ \mathbf{y}[k] &\approx \mathbf{g}(\bar{\mathbf{x}}[k], \mathbf{u}[k]) + \mathbf{G}_{\bar{\mathbf{x}}[k]} (\mathbf{x}[k] - \bar{\mathbf{x}}[k]) + \mathbf{v}[k]. \end{aligned} \right\} \quad (7)$$

Поскольку распределение шума и априорные распределения состояния при $k = 1$ — гауссовы, то

в линеаризованной системе (7) условное распределение вероятности состояния в произвольный момент времени при заданной последовательности входов и выходов также гауссово. Таким образом, сглаживатель Рауха может использоваться на линеаризованной системе (7) для вывода этого условного распределения. В противоположность линейному сглаживателю Рауха, в линеаризованном сглаживателе Рауха ошибка ковариации для оценки состояния и матрицы усиления Калмана зависит не только от наблюдений с текущим временным индексом.

На M -шаге (EM -алгоритма) проблема в том, что вектор-функции $\mathbf{f}(\cdot)$ и $\mathbf{g}(\cdot)$ обучаются, используя неопределенные оценки состояния посредством сглаживателя Рауха [5, 8]. Это затрудняет применение стандартных методов регрессии. Рассмотрим оценку $\mathbf{f}(\cdot)$ с параметрами $\mathbf{x}[k]$ и $\mathbf{u}[k]$ и результатом $\mathbf{x}[k+1]$. Для каждого k оцениваемое сглаживателем Рауха условное распределение является гауссовым с полной ковариацией в пространстве $\{\mathbf{x}[k], \mathbf{x}[k+1]\}$. Ниже разложения нелинейных функций $\mathbf{f}(\cdot)$ и $\mathbf{g}(\cdot)$ по набору гауссовых радиальных базисных функций (*radial basis functions* — RBFs) адаптируются к этим гауссовым выборкам данных.

Рассматривается представляющее функцию $\mathbf{f}(\cdot)$ нелинейное отображение векторов состояния \mathbf{x} и входа \mathbf{u} на вектор состояния \mathbf{z} :

$$\mathbf{z} = \sum_{i=1}^M \mathbf{h}[i] \rho_i(\mathbf{x}) + \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{b} + \mathbf{w}, \quad (8)$$

где \mathbf{w} — гауссов шум с нулевым средним и ковариацией \mathbf{Q} . Параметрами отображения (8) являются коэффициенты $\mathbf{h}[i]$, $i = \overline{1, M}$, при скалярных радиальных базисных функциях $\rho_i(\mathbf{x})$, $i = \overline{1, M}$, матрицы \mathbf{A} , \mathbf{B} и вектор смещения \mathbf{b} . Гауссовы радиальные базисные функции в пространстве векторов \mathbf{x} с центром $\mathbf{c}[i]$ и ковариационной матрицей $\mathbf{S}[i]$ имеют вид

$$\rho_i(\mathbf{x}) = |\mathbf{S}[i]|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{c}[i])^T \mathbf{S}[i]^{-1}(\mathbf{x} - \mathbf{c}[i])\right), \quad (9)$$

где $|\mathbf{S}[i]|$ — это детерминант матрицы $\mathbf{S}[i]$. Отображение (8) используется несколькими способами для представления систем (7), в зависимости от того, какое из отображений $\mathbf{f}(\cdot)$ или $\mathbf{g}(\cdot)$ считается нелинейным. Приведем три примера этого: 1) для $\mathbf{f}(\cdot)$ используются подстановки $\mathbf{x} \leftarrow \mathbf{x}[k]$, $\mathbf{u} \leftarrow \mathbf{u}[k]$ и $\mathbf{z} \leftarrow \mathbf{x}[k+1]$; 2) для $\mathbf{f}(\cdot)$ используются подстановки $\mathbf{x} \leftarrow (\mathbf{x}[k], \mathbf{u}[k])$, $\mathbf{u} \leftarrow \mathbf{0}$ и $\mathbf{z} \leftarrow \mathbf{x}[k+1]$; 3) для $\mathbf{g}(\cdot)$ используются подстановки $\mathbf{x} \leftarrow \mathbf{x}[k]$, $\mathbf{u} \leftarrow \mathbf{u}[k]$ и $\mathbf{z} \leftarrow \mathbf{y}[k]$. Поскольку набор данных для переменных $(\mathbf{x}, \mathbf{z}, \mathbf{u})$ является выборкой распределения в виде смеси гауссовых распределений, это распределение аналитически исключается посредством интегрирования, чтобы обеспечить соответствие RBF-модели (8). Распределение данных имеет вид

$$P(\mathbf{x}, \mathbf{z}, \mathbf{u}) = \frac{1}{J} \sum_j \mathcal{P}_j(\mathbf{x}, \mathbf{z}) \delta(\mathbf{u} - \mathbf{u}[j]), \quad (10)$$

где гауссово распределение $\mathcal{P}_j(\mathbf{x}, \mathbf{z}) \equiv \mathcal{N}(\mathbf{x}, \mathbf{z} | \boldsymbol{\mu}[j], \mathbf{C}[j])$ со средним $\boldsymbol{\mu}[j] = \{\boldsymbol{\mu}_x[j], \boldsymbol{\mu}_z[j]\}$ и ковариационной матрицей

$$\mathbf{C}[j] = \begin{bmatrix} \mathbf{C}_{xx}[j] & \mathbf{C}_{xz}[j] \\ \mathbf{C}_{zx}[j] & \mathbf{C}_{zz}[j] \end{bmatrix}.$$

Определим вектор

$$\bar{\mathbf{z}}_0(\mathbf{x}, \mathbf{u}) = \sum_{i=1}^M \mathbf{h}[i] \rho_i(\mathbf{x}) + \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{b} = \boldsymbol{\theta}\boldsymbol{\Phi},$$

где $\boldsymbol{\theta} = [\mathbf{h}[1], \dots, \mathbf{h}[M], \mathbf{A}, \mathbf{B}, \mathbf{b}]$ — вектор параметров и $\boldsymbol{\Phi} = [\rho_1(\mathbf{x}), \dots, \rho_M(\mathbf{x}), \mathbf{x}^T, \mathbf{u}^T, \mathbf{1}^T]^T$ — вектор переменных отображения (8). Поскольку в (8) шум \mathbf{w} — гауссов с нулевым средним и ковариационной матрицей \mathbf{Q} , то, в рамках RBF-модели, логарифм правдоподобия одного вектора \mathbf{z} имеет вид

$$-\frac{1}{2}(\mathbf{z} - \bar{\mathbf{z}}_0(\mathbf{x}, \mathbf{u}))^T \mathbf{Q}^{-1}(\mathbf{z} - \bar{\mathbf{z}}_0(\mathbf{x}, \mathbf{u})) - \frac{1}{2} \log |\mathbf{Q}| + c, \quad (11)$$

где c — константа. Поскольку набор $\{\mathbf{x}, \mathbf{z}\}$ — это гауссова выборка данных, максимум ожидаемого логарифма правдоподобия RBF-модели (8) дается минимизацией проинтегрированной квадратичной формы (11) (со знаком минус)

$$\min_{\boldsymbol{\theta}, \mathbf{Q}} \left\{ \sum_j \int_{\mathbf{z}} \mathcal{P}_j(\mathbf{x}, \mathbf{z}) (\mathbf{z} - \bar{\mathbf{z}}_0(\mathbf{x}, \mathbf{u}))^T \times \right. \\ \left. \times \mathbf{Q}^{-1}(\mathbf{z} - \bar{\mathbf{z}}_0(\mathbf{x}, \mathbf{u})) d\mathbf{x} d\mathbf{z} + J \log |\mathbf{Q}| \right\}. \quad (12)$$

Пусть угловые скобки $\langle \cdot \rangle_j$ обозначают ожидание по гауссову распределению \mathcal{P}_j (7). С учетом обозначений для вектора параметров $\boldsymbol{\theta}$ отображения (8) и вектора переменных $\boldsymbol{\Phi}$, и матричного тождества

$$(\mathbf{z} - \boldsymbol{\theta}\boldsymbol{\Phi})^T \mathbf{Q}^{-1}(\mathbf{z} - \boldsymbol{\theta}\boldsymbol{\Phi}) = \text{tr}(\mathbf{Q}^{-1}(\mathbf{z} - \boldsymbol{\theta}\boldsymbol{\Phi})(\mathbf{z} - \boldsymbol{\theta}\boldsymbol{\Phi})^T),$$

где $\text{tr}(\cdot)$ обозначает след (трек) матриц, формула (8) принимает следующий вид:

$$\min_{\boldsymbol{\theta}, \mathbf{Q}} \left\{ \text{tr} \left(\mathbf{Q}^{-1} \sum_j \langle (\mathbf{z} - \boldsymbol{\theta}\boldsymbol{\Phi})(\mathbf{z} - \boldsymbol{\theta}\boldsymbol{\Phi})^T \rangle_j \right) + J \log |\mathbf{Q}| \right\}. \quad (13)$$

Вычисляя в формуле (13) частную производную по $\boldsymbol{\theta}^T$ и приравнявая ее к нулю, получаем уравнение

$$\sum_j \langle (\mathbf{z} - \boldsymbol{\theta}\boldsymbol{\Phi})\boldsymbol{\Phi}^T \rangle_j = 0,$$

которое дает оценку

$$\boldsymbol{\theta} = \left(\sum_j \langle \mathbf{z}\boldsymbol{\Phi}^T \rangle_j \right) \left(\sum_j \langle \boldsymbol{\Phi}\boldsymbol{\Phi}^T \rangle_j \right)^{-1}. \quad (14)$$

Аналогично, выражение для оценки ковариационной матрицы \mathbf{Q} имеет вид

$$\mathbf{Q} = \frac{1}{J} \left(\sum_j \langle \mathbf{z}\mathbf{z}^T \rangle_j - \boldsymbol{\theta} \sum_j \langle \boldsymbol{\Phi}\mathbf{z}^T \rangle_j \right). \quad (15)$$

Ожидания, необходимые для оценки θ и Q по формулам (14), (15) следующие: $\langle \mathbf{x} \rangle_j$, $\langle \mathbf{z} \rangle_j$, $\langle \mathbf{x}\mathbf{x}^T \rangle_j$, $\langle \mathbf{z}\mathbf{z}^T \rangle_j$, $\langle \mathbf{x}\mathbf{z}^T \rangle_j$, $\langle \rho_i(\mathbf{x}) \rangle_j$, $\langle \mathbf{x}\rho_i(\mathbf{x}) \rangle_j$, $\langle \mathbf{z}\rho_i(\mathbf{x}) \rangle_j$ и $\langle \rho_i(\mathbf{x})\rho_j(\mathbf{x}) \rangle_j$, полные их выражения приведены в работах [5, 8]. Эти ожидания выражаются через параметры гауссовых распределений $\mathcal{P}_j(\mathbf{x}, \mathbf{z}) \equiv \mathcal{N}(\mathbf{x}, \mathbf{z} | \mu[j], \mathbf{C}[j])$ (10).

Практический успех *EM*-алгоритма зависит от его инициализации. Для нелинейных систем с линейной функцией выхода инициализация использует факторный анализ по максимуму правдоподобия, обучаемый на наборе наблюдений $\mathbf{y}[k]$, $k = \overline{1, N}$. Факторный анализ предполагает, что выходные переменные генерируются небольшим количеством независимых гауссовых скрытых состояний, и к каждой выходной переменной добавляется независимый гауссов шум [10–12]. Матрица весов (называемая матрицей нагрузок) обучается посредством факторного анализа для инициализации матрицы наблюдений \mathbf{G} динамической системы (7). Это дает оценки состояния в каждый момент времени. Эти оценки используются в нелинейной регрессии на основе RBF-функций. Перед инициализацией нелинейной регрессии проводится обучение линейной системы.

Эксперименты на многомерных временных рядах показывают возможности нелинейной байесовской фильтрации при использовании нелинейных динамических моделей.

2. Численный эксперимент

В настоящей работе анализ тонкой структуры кластеров траекторий движения самолетов, полученных методом полиномиальных регрессий [1, 2], выполняется с помощью адаптивного нелинейного фильтра Калмана, т. е. посредством модели нелинейных динамических систем (5).

В качестве анализируемых данных использовались траектории 117 самолетов, идущих на посадку в международном аэропорту и зарегистрированных радаром TRACON 1 января 2006 г. (<https://c3.nasa.gov/dashlink/resources/132/>). Начало координат совпадает с положением радара, интервал времени между точками регистрации составляет около 5 с. В работе учитываются только 160 последних точек каждой траектории, что позволяет исключить случайные маневры самолетов перед заходом на посадку. Эти траектории в трехмерном пространстве представлены в работе [2].

Пять кластеров траекторий самолетов (рис. 1, см. вторую сторону обложки) выделяются в результате применения метода полиномиальных регрессий [2]. Каждый кластер соответствует определенному "посадочному" паттерну или маршруту. Распределение числа траекторий по кластерам следующее: 16 траекторий в розовом кластере; 13 — в зеленом, 3, 37 и 38 — в синем, черном и красном кластерах соответственно.

На рис. 2 (см. вторую сторону обложки) показаны проекции траекторий в анализируемых кла-

стерах (см. рис. 1) на координатные оси x , y и z в соответствии с последовательностью моментов времени k регистрации радаром. Линии тренда, полученные в работе [2] методом полиномиальных регрессий и выделенные жирными линиями, представляют обобщенную форму траекторий в каждом кластере. Сходство траекторий движения самолетов в кластерах обусловлено существованием нескольких типичных маршрутов посадки ("посадочных" паттернов).

В работе [3] при использовании модели линейных динамических систем при анализе однородности кластеров в четырех из пяти рассматриваемых кластеров выделялась их тонкая (неоднородная) структура. Эти модели позволяют значительно сократить размерность анализируемого пространства, поэтому неоднородная структура кластеров выявляется при рассмотрении проекций траекторий каждого кластера на координатные оси x , y , z .

Выявление неоднородности в кластере (розом на рис. 1 и 2, см. вторую сторону обложки) с помощью метода нелинейных динамических систем демонстрируется на рис. 3 (см. вторую сторону обложки). Как видно на рис. 3, после определения скрытых состояний 1 и 2 (жирные линии красного и синего цвета, соответственно) для проекции на ось x траекторий розового кластера (см. рис. 2), этот кластер проекций траекторий разделяется на два подкластера: красный и синий. Приписывание траекторий к каждому из подкластеров 1 и 2 выполняется в соответствии с мерой косинуса между компонентами скрытого состояния $[x_i[k]$, $k = \overline{1, N}]$, $i = \overline{1, p}$, и компонентами проекции наблюдений (траекторий) $[y_j[k]$, $k = \overline{1, N}]$, $j = \overline{1, q}$,

$$R = \left(\sum_{k=1}^N x_i[k]y_j[k] \right) / \sqrt{\sum_{k=1}^N (x_i[k])^2} \sqrt{\sum_{k=1}^N (y_j[k])^2},$$

которая учитывает отличие в направлении векторов в пространстве состояний (безотносительно Евклидова расстояния между ними).

Сходство результатов анализа структуры кластеров траекторий, полученных в работе [3] с помощью линейной модели динамической системы и в настоящей статье с помощью нелинейной модели динамической системы, является обоснованием возможности применения в работе [3] линейной модели динамической системы к анализу кластеров траекторий движения воздушных судов, маршруты которых не вполне детерминированы.

Заключение

В работе [3] применение метода линейных динамических систем к анализу структуры кластеров посадочных траекторий самолетов основано на предположении о переходном характере проекций этих траекторий. В настоящей работе правомерность применения метода линейных динамических

систем к анализу структуры кластеров посадочных траекторий самолетов подтверждается подобием кластеров, полученных с помощью общих моделей нелинейных динамических систем, для тех же посадочных траекторий.

Список литературы

1. **Gaffney S., Smyth P.** Joint probabilistic curve clustering and alignment // Proc. of Neural Information Processing Systems (NIPS 2004). Advances in Neural Information Processing Systems / Eds. Saul L., Weiss Y., Bottou L. V. 17. Cambridge, MA: MIT Press. 2005. P. 473—480.
2. **Кухаренко Б. Г., Солнцева М. О.** Кластеризация управляемых объектов на основе сходства их многомерных траекторий // Информационные технологии. 2014. № 5. С.3—7.
3. **Кухаренко Б. Г., Солнцева М. О.** Анализ результатов кластеризации многомерных траекторий посредством моделей линейных динамических систем // Информационные технологии. 2015. Т. 21, № 2. С. 104—109.

4. **Shumway R. H., Stoffer D. S.** Time Series Analysis and Its Applications. New York: Springer, 2011.
5. **Roweis S., Ghahramani Z.** Learning nonlinear dynamical systems using the Expectation-Maximization algorithm / Ed. Haykin S. // Kalman Filtering and Neural Networks. John Wiley & Sons, 2001. P. 175—220.
6. **Simon D.** Optimal State Estimation: Kalman, H_∞ and Nonlinear Approaches. Hoboken, New Jersey: John Wiley & Sons, Inc, 2006.
7. **Einicke G. A.** Smoothing, Filtering and Prediction: Estimating the Past, Present and Future. Rijeka (Croatia): Intech. 2012.
8. **Ghahramani Z., Roweis S.** Learning nonlinear dynamical systems using an EM algorithm // Advances in Neural Information Processing Systems / Eds. Kearns M. S., Solla S. A., Cohn D. A. V. 11. Cambridge, MA: The MIT Press, 1999. P. 599—605.
9. **Kukharensko B. G.** Use of the Prony method for modal identification of slow-evolutionary linear structures // Journal of Structural Control. 2000. V. 7, N. 2. P. 203—218.
10. **Moody J., Darken C.** Fast learning in networks of locally-tuned processing units // Neural Computation. 1989. V. 1. P. 281—294.
11. **Broomhead D. S., Lowe D.** Multivariable functional interpolation and adaptive networks // Complex Systems. 1988. V. 2. P. 321—355.
12. **Roweis S., Ghahramani Z.** A unifying review of linear Gaussian models // Neural Computation. 1999. V. 11, N. 2. P. 305—345.

B. G. Kukharensko¹, M. O. Solntseva-Chalei²

¹Leading Research Scientist, Blagonravov Institute of Engineering Science of RAS,
e-mail: kukharenskobg@hotmail.com,

²Post-Graduate Student, Moscow Institute of Physics and Technology (SU),
e-mail: solntseva.chalei@gmail.com

Applying Nonlinear Dynamical System Models for Analysis of Multidimensional Trajectory Clustering Results

Problem of clustering trajectories is pre-conditioned by a need to organize motion of objects under control. Upon advisement the motion trajectories in three dimensional space, it is more naturally to select these into clusters according to the trajectory characteristic form. The polynomial regression method is the best approach to trajectory cluster selection, which estimates a form of general trajectory of each cluster. In present paper the analysis of trajectory clusters obtained by polynomial regression method is performed by means of adaptive nonlinear Kalman filter. In discrete time nonlinear Kalman filter, a nonlinear dynamical system describes state evolution on one time-step and current causality of state and observation. To estimate the nonlinear dynamical system unknown parameter and hidden states simultaneously, the Expectation maximization algorithm is applied, which use radial basis function approximation to express nonlinear dynamics of system state vector. Extracted components of system state vector gives an opportunity to elicit cluster fine structure in accordance with cosine measure.

Keywords: data mining, multi-dimensional trajectories, clustering, polynomial regression, Kalman filter, Rauch smoother, nonlinear dynamical systems, radial basis function, Expectation-Maximization algorithm

References

1. **Gaffney S., Smyth P.** Joint probabilistic curve clustering and alignment / Eds. Saul L., Weiss Y., Bottou L. *Proc. Neural Information Processing Systems (NIPS 2004)*. December 13—18, 2004. Vancouver, British Columbia, Canada. *Advances in Neural Information Processing Systems*. V. 17. Cambridge, MA: MIT Press. 2005. P. 473—480.
2. **Kukharensko B. G., Solntseva M. O.** Klasterizacia upravlyаемых объектов na osnove shodstva ih mnogomernykh trajektoriy. *Informacionnye tehnologii*. 2014. N. 5. P. 3—7.
3. **Kukharensko B. G., Solntseva M. O.** Analiz rezultatov klasterizacii mnogomernykh trajektoriy posredstvom modelei lineinykh dinamicheskikh sistem. *Informacionnye tehnologii*. 2015. V. 21, N. 2. P. 104—109.
4. **Shumway R. H., Stoffer D. S.** *Time Series Analysis and Its Applications*. New York: Springer, 2011.
5. **Roweis S., Ghahramani Z.** Learning nonlinear dynamical systems using the Expectation-Maximization algorithm / Haykin S., ed. *Kalman Filtering and Neural Networks*. John Wiley & Sons. 2001. P. 175—220.

6. **Simon D.** *Optimal State Estimation: Kalman, H_∞ and Nonlinear Approaches*. Hoboken, New Jersey: John Wiley & Sons, Inc, 2006.
7. **Einicke G. A.** *Smoothing, Filtering and Prediction: Estimating the Past, Present and Future*. Rijeka (Croatia): Intech. 2012.
8. **Ghahramani Z., Roweis S.** Learning nonlinear dynamical systems using an EM algorithm / Eds. Kearns M. S., Solla S. A., Cohn D. A. *Advances in Neural Information Processing Systems*. V. 11. Cambridge, MA: The MIT Press, 1999. P. 599—605.
9. **Kukharensko B. G.** Use of the Prony method for modal identification of slow-evolutionary linear structures. *Journal of Structural Control*. 2000. V. 7, N. 2. P. 203—218.
10. **Moody J., Darken C.** Fast learning in networks of locally-tuned processing units. *Neural Computation*. 1989. V. 1. P. 281—294.
11. **Broomhead D. S., Lowe D.** Multivariable functional interpolation and adaptive networks. *Complex Systems*. 1988. V. 2. P. 321—355.
12. **Roweis S., Ghahramani Z.** A unifying review of linear Gaussian models. *Neural Computation*. 1999. V. 11, N. 2. P. 305—345.

И. В. Мишин, д-р физ.-мат. наук, ст. науч. сотр., проф., e-mail: i.v.mishin@mail.ru
"МАТИ"—Российский государственный технологический университет им. К. Э. Циолковского

Численное моделирование однолинейных систем обслуживания

Построены численные алгоритмы для расчета выходных характеристик систем типа $GI|GI|1$ в стационарном режиме работы. Алгоритмы реализованы с помощью метода статистических испытаний. Рассматриваются эмпирические распределения времени ожидания запроса в очереди и времени между моментами окончания обслуживания соседних запросов. Вычисляются значения статистических оценок средних и средних квадратических отклонений времени ожидания и времени задержки. Построена автокорреляционная функция реализаций процесса ожидания. Вычислительная программа тестирована путем сравнения эмпирических и теоретических значений средних времени ожидания в случае $M|E|1$. Выполнен сравнительный количественный анализ систем $M|E|1$, $E|E|1$ и $GI|GI|1$. Исследованы зависимости длины очереди и радиуса корреляции последовательности интервалов ожидания от нагрузки системы и дисперсии интервалов времени между моментами поступления запросов.

Ключевые слова: однолинейная система обслуживания, запрос, нагрузка, очередь, время ожидания, время обслуживания, время задержки, эмпирическое распределение, статистическая оценка

Введение

Аналитические методы теории массового обслуживания [1, 2] позволяют рассчитывать выходные характеристики стохастических систем с очередями в предположении о марковских или полумарковских свойствах этих систем. Для имитации работы систем с любыми допустимыми потоками входных запросов и обслуживания необходимо располагать универсальными численными алгоритмами и соответствующим программным обеспечением. В общем случае численные алгоритмы для описания систем обслуживания основываются на известных рекуррентных соотношениях, связанных с последовательностью запросов [1–6] и реализуются с помощью метода статистических испытаний [4, 7]. Некоторые программные средства для расчета систем обслуживания указаны в работах [7, 8]. Наиболее известна среда GPSS Word [8, 9]. Положительным качеством GPSS Word является относительная простота использования. В то же время это приложение не обеспечивает высокую точность вычислений и не позволяет проводить подробные расчеты параметров моделируемых систем. Мощным средством создания вычислительных программ для расчета стохастических систем являются языки программирования высокого уровня. Однако их использование в исследовательских целях является самостоятельной трудоемкой задачей.

Реализация метода статистических испытаний значительно упростилась с появлением специальных вычислительных средств, обладающих датчиками псевдослучайных чисел с заданными законами распределения и приемлемой вычислительной погрешностью. Ниже рассматривается алгоритм численного моделирования стационарного распределения времени пребывания запроса в однолинейной системе общего вида с дисциплиной обслуживания

FIFO. Для реализации алгоритма использовалась вычислительная система MathCAD [10].

В работе используются следующие обозначения: $n + 1$ — число запросов; $\{\zeta_i\}_{0, n}$ — моменты поступления запросов в систему; $\tau_i = \zeta_i - \zeta_{i-1}$ — время между поступлениями запросов (длительности входных интервалов), $i = \overline{1, n}$; $a(t)$, $A(t)$, $\bar{\tau}$, σ_τ , c_τ — плотность, функция распределения, среднее, среднее квадратическое отклонение (СКО) и коэффициент вариации длительности входных интервалов; v_i — время обслуживания i -го запроса, $i = \overline{0, n}$; $b(t)$, $B(t)$, \bar{v} , σ_v , c_v — плотность, функция распределения, среднее, СКО и коэффициент вариации времени обслуживания одного запроса; w_i — время пребывания i -го запроса в очереди (время ожидания), $i = \overline{0, n}$; \bar{w} — среднее время ожидания; \bar{T} — среднее время пребывания запроса в системе (время задержки); $\{\eta_i\}_{0, n}$ — моменты окончания обслуживания запросов; $z_j = \eta_{j+1} - \eta_j$ — длительности выходных интервалов, $j = \overline{0, n-1}$; $\rho = \bar{v}/\bar{\tau}$ — нагрузка. Время поступления первого запроса ζ_0 — начало отсчета времени. Прочие обозначения вводим по ходу изложения. Реализации случайных последовательностей $\{\tau_i\}_{1, n}$ и $\{v_i\}_{0, n}$ генерируются датчиками псевдослучайных чисел с плотностями распределения $a(t)$ и $b(t)$ соответственно. Существование стационарного режима функционирования системы обслуживания гарантирует условие $\rho < 1$.

Вычислительный алгоритм

Поскольку первый запрос поступает в пустую систему, полагаем $w_0 = 0$. Второй запрос находит

обслуживающее устройство свободным при $\tau_1 \geq v_0$ и $w_1 = 0$. При $\tau_1 < v_0$ устройство занято, и второй запрос находится в очереди в течение времени $w_1 = v_0 - \tau_1$. Таким образом, $w_1 = \max[v_0 - \tau_1, 0]$. Следуя [1, Гл. 8], введем величину

$$u_i = v_i - \tau_{i+1}, i = \overline{0, n-1}. \quad (1)$$

Для всех $i \geq 1$ имеет место рекуррентное соотношение

$$w_{i+1} = \max[w_i + u_i, 0]. \quad (2)$$

Формула (2) с учетом $w_0 = 0$ определяет процедуру расчета значений $\{w_j\}_{0,n}$.

Рассмотрим теперь последовательность моментов окончания обслуживания запросов. Время окончания обслуживания первого запроса наступает в момент $\eta_0 = v_0$. Для второго запроса при $\zeta_1 \geq v_0$ ($\tau_1 \geq v_0$, $u_0 \geq 0$) момент окончания обслуживания определяется равенством $\eta_1 = \tau_1 + v_1$. Если же $\zeta_1 < v_0$, ($\tau_1 < v_0$, $u_0 < 0$), то $\eta_1 = v_0 + v_1 = \eta_0 + v_1$. Объединяя оба случая, получаем $\eta_1 = \max[\eta_0, \zeta_1] + v_1$. Для всех $j \geq 1$ имеем $\eta_j = \zeta_j + v_j$ при $\zeta_j \geq \eta_{j-1}$, и $\eta_j = \eta_{j-1} + v_j$ при $\zeta_j < \eta_{j-1}$. Следовательно,

$$\eta_j = \max[\eta_{j-1}, \zeta_j] + v_j, j \geq 1. \quad (3)$$

Формула (3) с учетом $\eta_0 = v_0$ определяет моменты выхода запросов из системы. Далее вычисляем длительности выходных интервалов:

$$z_j = \eta_{j+1} - \eta_j, j = \overline{0, n-1}. \quad (4)$$

Формулы (1)–(4) составляют основу расчетного алгоритма. Стационарный режим в системе устанавливается постепенно. Поэтому в расчетах следует исключить из рассмотрения некоторую начальную часть потока запросов — отрезок нестационарности.

Располагая реализациями случайных последовательностей $\{\tau_j\}_{1,n}$, $\{v_j\}_{0,n}$, $\{w_j\}_{0,n}$, $\{z_j\}_{0,n-1}$, можно вычислить все числовые характеристики системы обслуживания с точностью, которая определяется числом n и ошибками округления. Вычислялись следующие параметры и функции: $\hat{\tau}$, $\hat{\sigma}_\tau$, \hat{c}_τ (реализации статистических оценок будем обозначать теми же символами) — выборочные среднее, СКО и коэффициент вариации длительности входных интервалов; \hat{w} , $\hat{\sigma}_w$, \hat{c}_v — выборочные среднее, СКО и коэффициент вариации времени ожидания; \hat{T} , $\hat{\sigma}_T$, \hat{c}_T — выборочные среднее, СКО и коэффициент вариации времени задержки; $N_0, \gamma_0 = N_0/N$ — число и доля запросов с нулевым временем ожидания в стационарной очереди, $N = n + 1 - n_0$, n_0 — длина отрезка нестационарности; $\{q_j\}_{0,N-N_0}$, $h_q(t)$, $\hat{Q}(t)$ — эмпирическое распределение, гистограмма и функция распределения запросов с ненулевым временем

ожидания; $h_z(t)$, $\hat{Z}(t)$ — гистограмма и эмпирическая функция распределения длительности выходных интервалов; $\{R_k\}_{0,M}$ — дискретная автокорреляционная функция (набор коэффициентов корреляции времени ожидания запросов, отстоящих друг от друга на k номеров), $M < [N/4]$.

При выборе числа испытаний n принимались во внимание следующие соображения. Случайная последовательность $\{\tau_j\}_{1,n}$ подчиняется закону больших чисел и к ней применима центральная предельная теорема [11]. С уровнем доверия 0,997 можно утверждать, что $P\{|\hat{\tau} - \bar{\tau}| \leq 3\sigma_\tau/\sqrt{n}\} \approx 0,997$.

Здесь $\hat{\tau} = \sum_{i=1}^n x_i$ — значение выборочного среднего,

$\{x_j\}_{1,n}$ — реализация случайной последовательности $\{\tau_j\}_{1,n}$. Потребуем, чтобы относительная погрешность $\delta_\tau = |\hat{\tau} - \bar{\tau}|/\bar{\tau}$ не превосходила определенное значение: $\delta \leq 0,001$. Число n найдем из соотношения

$\delta = 3\sigma_\tau/\bar{\tau}\sqrt{n}$; $n = (3\sigma_\tau/\bar{\tau}\delta)^2$. С параметрами простейшего входного потока $\lambda = 1$, $\bar{\tau} = \sigma_\tau = \lambda^{-1} = 1$

и $\delta = 10^{-3}$ имеем $n = 9 \cdot 10^6$. Для последовательности $\{v_j\}_{1,n}$ в силу неравенства $\bar{v} < \bar{\tau}$ и при условии $\sigma_v < \sigma_\tau$ число n примет меньшее значение. Как показывают численные эксперименты, при заданном n реальное значение относительной погрешности δ оказывается значительно меньше определенного выше. Принимая во внимание это обстоятельство и учитывая ограничение на размеры массивов в системе MathCAD, в расчетах полагаем $n = 10^6$.

Для выяснения репрезентативности исходных генераций псевдослучайных чисел эмпирические плотности $\hat{a}(t)$, $\hat{b}(t)$ и значения случайных величин $\hat{\tau}$, $\hat{\sigma}_\tau$, \hat{v} , $\hat{\sigma}_v$ сравнивались с исходными аналитическими плотностями $a(t)$, $b(t)$ и теоретическими значениями величин $\bar{\tau}$, σ_τ , \bar{v} , σ_v соответственно. Для получения наиболее объективных статистических оценок проводилось усреднение выборочных

характеристик. Например, $(\hat{\tau})_p = \frac{1}{p} \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n x_i^{(p)}$,

$(\hat{\sigma}^2)_p = \frac{1}{p} \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n (x_i^{(p)} - \hat{\tau})^2$, где $x_i^{(p)}$ — значение

случайной величины τ_i в p -м испытании. Число испытаний полагалось равным $p = 30$. Аналогично определялись усредненные точечные статистические оценки остальных исследуемых величин. Процедуры

построения гистограмм и эмпирических функций распределения здесь не приводим.

Дискретная функция автокорреляции времени ожидания рассчитывалась по формуле

$$\hat{R}_k = \frac{1}{(L-k)\hat{d}} \sum_{i=0}^{L-k-1} (q_i - \hat{w})(q_{i+k} - \hat{w}), \quad (5)$$

где $\hat{d}_k = \frac{1}{L} \sum_{i=0}^{L-1} (q_i - \hat{w})^2$, $k = \overline{0, M}$, $M \ll L < n$.

Функция (5) приближается экспонентой $K(y = e^{-\omega y})$ с декрементом затухания ω . Тестирование алгоритма проводили путем сравнения значений оценок \hat{w} , $\hat{\sigma}_w$ с теоретическими значениями величин \bar{w} , σ_w в системе $M|E|1$. Формулы для расчета теоретических значений имеют вид [1]

$$\bar{w} = \frac{\rho \bar{v}(1 + c_v^2)}{2(1 - \rho)}, \quad (6)$$

$$\sigma_w = \sqrt{\bar{w}^2 + \frac{\rho v^3}{3\bar{v}(1 - \rho)}}, \quad (7)$$

где $v^3 = \int_0^\infty t^3 b(t) dt$ — третий начальный момент рас-

пределения времени обслуживания. Нетрудно заметить, что $\sigma_w > \bar{w}$. Для эрланговского обслуживания с масштабным параметром μ параметром формы r третий начальный момент равен $v^3 = \mu^{-3}r(r+1)(r+2)$.

Численные расчеты

Расчеты выполнялись для различных вариантов, обозначаемых кодировкой $GI|GI|1$. В настоящем анализе ограничимся случаями $M|E|1$, $E|E|1$, $GI|E|1$. Время обслуживания задаем распределением Эрланга с масштабным параметром $\mu = 8$ и параметром формы $r = 4$. При этом $\bar{v} = r\mu^{-1} = 0,5$, $\sigma_v = \sqrt{r} \mu^{-1} = 0,25$, $c_v = 0,5$, $\bar{w} = 5\rho/16(1 - \rho)$, $\sigma_w = \sqrt{\bar{w}(\bar{w} + 1/2)}$. В случае $GI|E|1$ длительности входных интервалов задавались распределением Вейбулла—Гнеденко. Значения параметров распределения длительности входных интервалов рассчитывались в зависимости от нагрузки ρ . Представленные ниже численные результаты получены для значений нагрузки $\rho = 0,5$ и $\rho = 0,8$.

1⁰. $M|E|1$. Пусть λ — интенсивность простейшего входного потока запросов. Учитывая, что $\bar{\tau} = \lambda^{-1}$, $\bar{v} = r\mu^{-1} = 0,5$, $\rho = \bar{v}/\bar{\tau}$, имеем $\lambda = \rho(\mu/r)$. Это соотношение при фиксированных μ , r связывает λ и ρ линейной зависимостью. При изменении ρ в интервале $[0, 1)$ величина λ меняется в интервале $[0, \bar{v}^{-1}) = [0, \mu/r)$. Рис. 1, а и б иллюстрирует точность исходных генераций псевдослучайных чисел при $n = 10^6$.

В табл. 1 приводятся усредненные выборочные средние $(\hat{\tau})_p$, $(\hat{v})_p$ и их усредненные погрешности $(\Delta_\tau)_p = (|\hat{\tau} - \bar{\tau}|)_p$, $(\delta_\tau)_p = (\Delta_\tau/\bar{\tau})_p$, $(\Delta_v)_p = (|\hat{v} - \bar{v}|)_p$, $(\delta_v)_p = (\Delta_v/\bar{v})_p$. Усредненные выборочные СКО составили $(\hat{\sigma}_\tau)_p = 1,00061$ и $(\hat{\sigma}_v)_p = 1,0006$. Усредненные погрешности оценок средних и СКО не превосходят 0,1 %. Поскольку разброс погрешностей лежит в диапазоне $\sim 10^{-5} \dots 10^{-3}$, для отдельных генераций погрешность снижается до $\sim 10^{-5}$.

На рис. 2 показано время ожидания 25 смежных запросов в стационарной очереди. В расчетах полагалось $n_0 = 10^3$.

Как видно на рис. 2, процесс образования очереди время от времени возвращается к состоянию $w = 0$, т. е. является регенерирующим. В табл. 2 сравниваются значения теоретических и усредненных выборочных средних времени ожидания w и времени задержки T . Усредненные абсолютные и относительные погрешности определяются формулами $(\Delta_w)_p = (|\hat{w} - \bar{w}|)_p$, $(\delta_w)_p = (\Delta_w/\bar{w})_p$, $(\Delta_T)_p = (|\hat{T} - \bar{T}|)_p$, $(\delta_T)_p = (\Delta_T/\bar{T})_p$. Так как $T = w + v$, то $\bar{T} = \bar{w} + \bar{v}$, $\hat{T} = \hat{w} + \hat{v}$. По условию задачи случайная величина v не зависит от w . Следовательно,

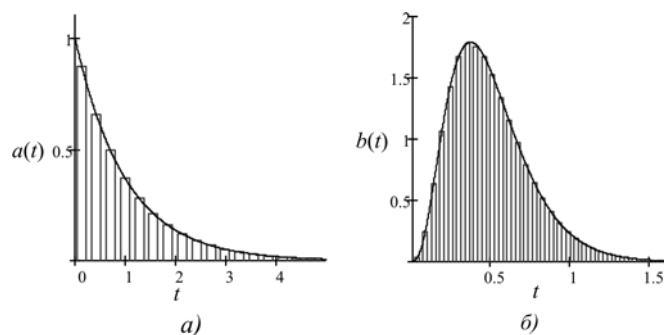


Рис. 1. Гистограмма длительностей входных интервалов простейшего потока (а); огибающая — теоретическая плотность $a(t)$; $\tau = 1$; гистограмма времени обслуживания (б); огибающая — теоретическая плотность $b(t)$; $\bar{v} = 0,5$

Таблица 1

Усредненные выборочные средние и их погрешности

$(\hat{\tau})_p$	$(\Delta_\tau)_p$	$(\delta_\tau)_p$	$(\hat{v})_p$	$(\Delta_v)_p$	$(\delta_v)_p$
1,00021	0,00091	0,00091	0,50001	0,00024	0,00052

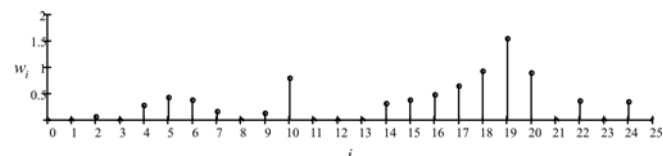


Рис. 2. Время ожидания в зависимости от номера запроса; $\rho = 0,5$

Теоретические значения среднего времени ожидания и среднего времени задержки и их усредненные статистические оценки для M/E1

ρ	\bar{w}	$(\hat{w})_p$	$(\Delta_w)_p$	$(\delta_w)_p$	\bar{T}	$(\hat{T})_p$	$(\Delta_T)_p$	$(\delta_T)_p$
0,5	0,3125	0,31246	0,00098	0,0028	0,8125	0,81247	0,00111	0,00135
0,8	1,25	1,24624	0,01103	0,0088	1,75	1,74619	0,01113	0,00767

Теоретические значения СКО и коэффициентов вариации времени ожидания и времени задержки и их усредненные статистические оценки для M/E1

ρ	σ_w	$(\hat{\sigma}_w)_p$	c_w	$(\hat{c}_w)_p$	σ_T	$(\hat{\sigma}_T)_p$	c_T	$(\hat{c}_T)_p$
0,5	0,50389	0,50362	1,61	1,61	0,5625	0,56265	0,69	0,69
0,8	1,47902	1,47262	1,18	1,18	1,4995	1,49369	0,86	0,86

$\sigma_T = \sqrt{\sigma_w^2 + r\mu^{-2}}$, $\hat{\sigma}_T = \sqrt{\hat{\sigma}_w^2 + \hat{\sigma}_v^2}$. Для $\rho = 0,8$ входные параметры имеют следующие значения: $\lambda = 1,6$, $\bar{\tau} = \sigma_v = \bar{v}/\rho = 0,625$.

При $\rho = 0,5$ значения погрешностей из табл. 2 превышают аналогичные значения из табл. 1, но не выходят за пределы $\sim 10^{-3}$. При возрастании нагрузки до $\rho = 0,8$ абсолютная погрешность в среднем возрастает почти на порядок и составляет доли процента, причем относительная погрешность возрастает приблизительно на полпорядка. Среднее число заявок в очереди можно определить по формуле Литтла: $\bar{m} = \lambda \bar{w}$ ($\hat{m} = \lambda \hat{w}$). При нагрузке $\rho = 0,5$ и $\rho = 0,8$ получаем соответственно $\bar{m} = 0,3125$ и $\bar{m} = 1,25$.

В табл. 3 сравниваются значения теоретических и усредненных выборочных СКО и коэффициентов вариации времени ожидания w и времени задержки T .

Как следует из табл. 2 и 3 погрешности определения средних величин и СКО приблизительно равны.

На рис. 3 представлена гистограмма ненулевого времени ожидания.

Доля запросов с нулевым временем ожидания связана с нагрузкой линейной зависимостью $\gamma_0 = 1 - \rho$.

На рис. 4 приведены графики эмпирической плотности длительности выходных интервалов и аналитических плотностей длительности входных интервалов и времени обслуживания.

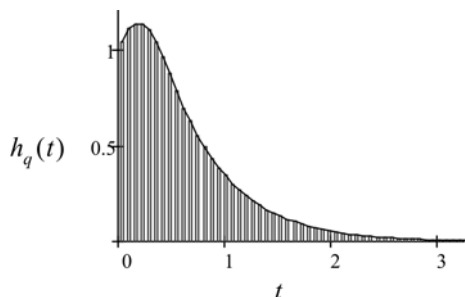


Рис. 3. Гистограмма ненулевого времени ожидания; $\gamma_0 = 0,5$; сплошной линией показана огибающая гистограммы

На рис. 5 приведены графики функций распределения длительности входных и выходных интервалов и времени ожидания в очереди и задержки в системе.

На рис. 6 показана автокорреляционная функция эмпирического распределения времени ожидания и ее аппроксимация.

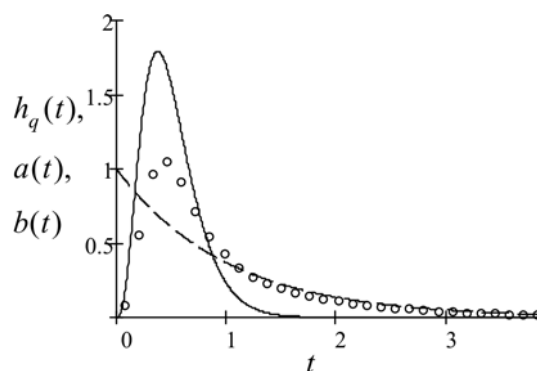


Рис. 4. Эмпирическая плотность длительности выходных интервалов (кружки) и аналитические плотности длительности входных интервалов (штриховая линия) и времени обслуживания (сплошная линия); $\rho = 0,5$

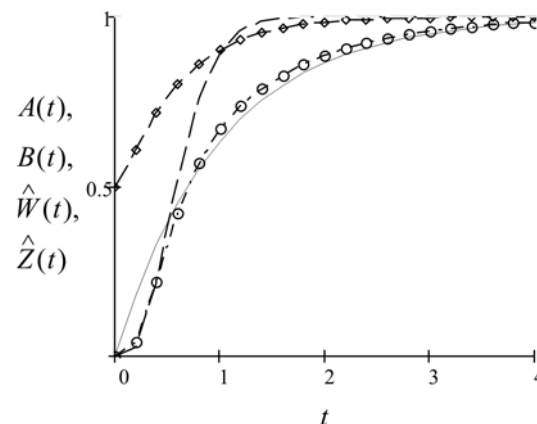


Рис. 5. Функции распределения длительности входных интервалов (сплошная линия), времени обслуживания (штриховая линия), времени ожидания (ромбики) и длительности выходных интервалов (кружки); $\rho = 0,5$, $\hat{W}(0) = \gamma_0$

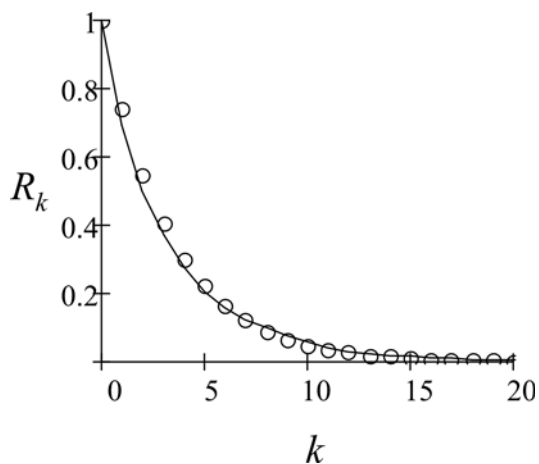


Рис. 6. Эмпирическая функция автокорреляции времени ожидания (кружки) и ее аппроксимация (сплошная линия), $\rho = 0,5$, $\omega = 0,3$

Отрезок заметной корреляции (радиус корреляции) для каждой реализации определяем равенством $e^{-\omega y} = 0,1$, откуда $y_0 = \omega^{-1} \ln 10$ и $k_0 = [y_0]$. В данном примере на отрезке заметной корреляции укладывается $k_0 = 7$ запросов. При $\rho = 0,8$ та же величина принимает значение $k_0 = 57$, а значение декремента затухания $\omega = 0,04$. Автокорреляционная функция сильно зависит от ρ . При малых ρ корреляция практически не наблюдается. При $\rho \rightarrow 1$ радиус корреляции неограниченно возрастает.

Заметим, что последовательность длительностей входных интервалов, интервалов обслуживания и выходных интервалов являются δ -коррелированными. При $r = 1$ система $M|E|1$ приобретет частный вид $M|M|1$. В этом случае для любых $\rho < 1$ справедлива теорема Берке. Сравнение выходных характеристик системы $M|E|1$, рассчитанных аналитически и методом статистических испытаний, тестируют вычислительную программу.

2⁰. $E|E|1$. Параметры обслуживания оставляем те же. Параметр формы распределения длительности входных интервалов полагаем равным $\nu = 4$. При $\rho = 0,5$ последовательно определяем среднюю длительность входных интервалов $\bar{\tau} = \bar{v}/\rho = 1$, масштабный множитель входного потока $\lambda = \nu \bar{\tau}^{-1} = 4$, СКО длительностей входных интервалов $\sigma_\tau = \sqrt{\nu} \lambda^{-1} = 0,5$ и коэффициент вариации $c_\tau = \sigma_\tau / \bar{\tau} = 0,5$. При $\rho = 0,8$ аналогично получаем $\bar{\tau} = 0,625$, $\lambda = 6,4$, $\sigma_\tau = 0,3125$, $c_\tau = 0,5$. В табл. 4 приведены усредненные выборочные значения среднего, СКО и коэффициента вариации времени ожидания и времени задержки.

Значение величины γ_0 , радиус корреляции времени ожидания в единицах запросов k_0 и декремент затухания ω эмпирической автокорреляционной функции времени ожидания равны соответственно $\gamma_0 = 0,78$, $k_0 = 3$, $\omega = 0,7$ при $\rho = 0,5$ и $\gamma_0 = 0,37$,

$k_0 = 20$, $\omega = 0,11$ при $\rho = 0,8$. Усреднение по числу генераций эти значения не меняет.

3⁰. $GI|E|1$. Сохраняем параметры обслуживания п. 1⁰. Длительности входных интервалов подчиняются распределению Вейбулла—Гнеденко с параметром формы α . В этом случае $\tau = \bar{v}/\rho = 0,5/\rho$, $\lambda = [\bar{\tau}/\Gamma(\alpha^{-1} + 1)]^{1/\alpha}$ — масштабный множитель, $\sigma_\tau = \lambda^{-1/\alpha} [\Gamma(2\alpha^{-1} + 1) + \Gamma^2(\alpha^{-1} + 1)]^{1/2}$. В расчетах полагаем $\alpha = 3$. В табл. 5 сведены значения оценок $(\hat{w})_\rho$, $(\hat{\sigma}_w)_\rho$, $(\hat{T})_\rho$, $(\hat{\sigma}_T)_\rho$, \hat{c}_w , \hat{c}_T . При $\rho = 0,5$ имеем $\bar{\tau} = 1$, $\lambda = 0,712$, $\sigma_\tau = 0,363$, $\bar{\gamma}_0 = 0,84$, $k_0 = 2$, $\omega = 0,95$; при $\rho = 0,8$: $\bar{\tau} = 0,625$, $\lambda = 2,917$, $\sigma_\tau = 0,227$, $\bar{\gamma}_0 = 0,43$, $k_0 = 17$, $\omega = 0,13$.

Из сравнения табл. 3, 4 и 5 можно заключить, что длина очереди сильно зависит от σ_τ . Наибольший разброс длительностей входных интервалов имеет место в простейшем потоке. Двойное снижение σ_τ при переходе от $M|E|1$ к $E|E|1$ приводит к уменьшению \bar{w} в 5 раз при $\rho = 0,5$ и в 3 раза при

Таблица 4

Усредненные статистические оценки среднего, СКО и коэффициента вариации времени ожидания и времени задержки для $E|E|1$

ρ	$(\hat{w})_\rho$	$(\hat{\sigma}_w)_\rho$	$(\hat{T})_\rho$	$(\hat{\sigma}_T)_\rho$	\hat{c}_w	\hat{c}_T
0,5	0,0649	0,17653	0,56491	0,30603	8,7	0,54
0,8	0,41869	0,59589	0,91874	0,64621	1,42	0,7

Таблица 5

Усредненные статистические оценки среднего, СКО и коэффициента вариации времени ожидания и времени задержки для $GI|E|1$

ρ	$(\hat{w})_\rho$	$(\hat{\sigma}_w)_\rho$	$(\hat{T})_\rho$	$(\hat{\sigma}_T)_\rho$	\hat{c}_w	\hat{c}_T
0,5	0,04158	0,13640	0,54160	0,28479	3,28	0,53
0,8	0,30266	0,46655	0,80279	0,52931	1,54	0,66

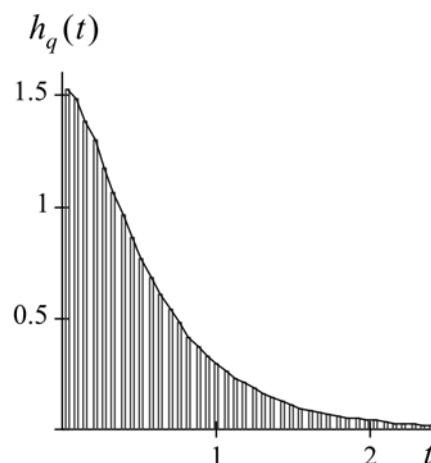


Рис. 7. Гистограмма ненулевого времени ожидания; $\rho = 0,8$

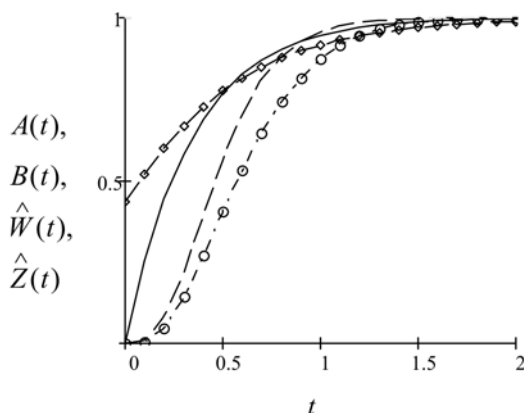


Рис. 8. Функции распределения длительности входных интервалов (сплошная), времени обслуживания (штриховая линия), времени ожидания (ромбики) и длительности выходных интервалов (кружки); $\rho = 0,8$, $\hat{W}(0) = \gamma_0$

$\rho = 0,8$. При переходе от $M|E|1$ к $GI|E|1$ величина \bar{w} уменьшается в 7 и 4 раза соответственно.

Форма гистограммы $h_q(t)$ и форма кривых $\hat{W}(t)$, $\hat{Z}(t)$ меняются при смене моделей. Иллюстрацией служат рис. 7 и 8.

Заключение

Статистическое моделирование позволяет построить достаточно универсальную схему для численного исследования однолинейных систем обслужи-

вания. Точность моделирования определяется датчиками псевдослучайных чисел, имитирующих поток входных запросов и время обслуживания. Представленные результаты можно обобщить на многолинейные системы и системы с фазовым обслуживанием. Рассмотренные алгоритмы применимы в задачах с самыми разнообразными последовательностями запросов. Возможен учет временной изменчивости интенсивности входного потока.

Список литературы

1. Крейпрок Л. М. Теория массового обслуживания. М.: Машиностроение, 1979.
2. Cohen J. W. The single server queue. Amsterdam [etc.]: North-Holland, 1992.
3. Климов Г. П. Стохастические системы обслуживания. М.: Наука, 1966.
4. Лифшиц А. Л., Мальц Э. А. Статистическое моделирование систем массового обслуживания. М.: Сов. Радио, 1978.
5. Калашников В. В., Рачев С. Т. Математические методы построения стохастических моделей обслуживания. М.: Наука, 1988.
6. Гнененко Б. В., Коваленко И. Н. Введение в теорию массового обслуживания. М.: Изд-во ЛКИ, 2007.
7. Cooper R. B. Introduction to Queueing Theory. London: Edward Arnold, 1981.
8. Крылов В. В., Самохвалова С. С. Теория телетрафика и ее приложения. СПб.: БХВ—Петербург, 2005.
9. Кудрявцев Е. М. GPSS Word. Основы имитационного моделирования различных систем. М.: ДМК Пресс, 2004.
10. Кирьянов Д. В. Mathcad 14. СПб.: БХВ—Петербург, 2007.
11. Соболев И. М. Численные методы Монте-Карло. М.: Наука, 1973.

I. V. Mishin, Professor, e-mail: i.v.mishin@mail.ru,

MATI — Russian State Technological University named after K. E. Tsiolkovsky

Numerical Modeling of Single-Server Systems

The numerical algorithms for calculation of output characteristics of the $GI|GI|1$ systems in stationary operation mode have been created. The algorithms have been realized by statistic testing method. Empirical distributions of waiting time of a request in the queue and time intervals between the moments of serving finishing of adjacent requests are considered. Values of the statistic estimations for mean and standard deviations of waiting time and holding time are calculated. The autocorrelation function of waiting process realizations has been created. Computer program has been tested by means of the comparison of the quantities of empirical mean with the theoretical quantities in case $M|E|1$. Comparative quantitative analysis of the system $M|E|1$, $E|E|1$ and $GI|E|1$ is performed. The dependences of queue length and correlation radius of waiting intervals succession from traffic intensity and dispersion of time intervals between the moments of the arrival of requests are researched.

Keywords: single-server system, request, traffic intensity, queue, waiting time, service time, holding time, empirical distribution, statistical estimation

References

1. Kleinrock L. M. Teorija massovogo obsluzhivaniya. M.: Mashinostroenie, 1979.
2. Cohen J. W. The single server queue. Amsterdam [etc.]: North-Holland, 1992.
3. Klimov G. P. Stokhasticheskie sistemy obsluzhivaniya. M.: Nauka, 1966.
4. Lifshic A. L., Mal'c Je. A. Statisticheskoe modelirovanie sistem massovogo obsluzhivaniya. M.: Sov. Radio, 1978.
5. Kalashnikov V. V., Rachev S. T. Matematicheskie metody postroeniya stokhasticheskikh modelej obsluzhivaniya. M.: Nauka, 1988.
6. Gnedenko B. V., Kovalenko I. N. Vvedenie v teoriyu obsluzhivaniya. M.: Izd-vo LKI, 2007.
7. Cooper R. B. Introduction to Queueing Theory. London: Edward Arnold, 1981.
8. Krylov V. V., Samohvalova S. S. Teorija teletrafika i ee prilozheniya. SPb.: BHV—Peterburg, 2005.
9. Kudrjavcev E. M. GPSS Word. Osnovy imitacionnogo modelirovaniya razlichnykh sistem. M.: DMK Press, 2004.
10. Kir'janov D. V. Mathcad 14. SPb.: BHV—Peterburg, 2007.
11. Sobol' I. M. Chislennyye metody Monte-Karlo. M.: Nauka, 1973.

Ускорение высокоточных вычислений за счет распараллеливания операции округления в комплексе систем счисления¹

Предложен способ ускорения высокоточных вычислений за счет распараллеливания операции округления на основе комплекса систем счисления: модулярной и знакоразрядной. Представлены преимущества данного подхода при выполнении других немодульных операций, таких как обратное преобразование в позиционную систему счисления, сравнение чисел, определение знака.

Ключевые слова: модулярная арифметика, высокоточные вычисления, избыточная система

Введение

В вычислительной практике при решении ряда прикладных задач на ЭВМ в таких областях как нанoeлектроника, ядерная физика, робототехника и др. требуются высокоточные компьютерные вычисления. Библиотеки, их поддерживающие, такие как, например, ZREAL (Россия), MPARITH (Германия), GMP (США) и др., имеют недостаток — резкое возрастание времени вычислений при увеличении числа арифметических операций или точности [1]. Данный недостаток препятствует их применению в суперкомпьютерных вычислениях и других областях, критичных к скорости выполнения арифметических операций.

Известны следующие достоинства и недостатки модулярной и симметричной знакоразрядной систем счисления [2, 3]:

- высокая скорость выполнения умножения, сложения, вычитания в модулярной системе счисления;
- сложность выполнения немодульных операций: деления, сравнения, округления, определение знака и др. в модулярной системе счисления;
- отсутствие распространения переносов при сложении в симметричной знакоразрядной системе счисления.

В работе [4] представлен подход к ускорению высокоточных вычислений за счет применения модулярной системы счисления (МСС), в которой получен эффект ускорения при решении ряда частных задач.

В высокоточных вычислениях в модулярной арифметике по предложенной схеме узким звеном, снижающим эффект ускорения, является операция

округления. Дальнейшие исследования посвящены ускорению операции округления [5].

Цель настоящей работы состоит в ускорении операции округления за счет использования достоинств модулярной и знакоразрядной систем счисления, т. е. комплекса систем счисления.

Рассмотрим формат представления чисел, алгоритмы выполнения операций в этом формате, способ обнаружения необходимости округления и округления на основе комплекса систем счисления.

1. Формат представления чисел в МСС

Пусть

$$A = K \cdot 10^t, \quad (1)$$

где K — целое число, такое, что $|K| \leq 10^{n_f} - 1$, $t \geq 0$ — порядок, такой, что удовлетворяет неравенству $0 \leq t \leq k_f$; n_f — натуральное число, характеризующее длину мантииссы числа с плавающей точкой; k_f — целое число, характеризующее диапазон представимых чисел.

Модулярный формат представления чисел вида (1), предложенный в работе [4], имеет вид:

$$A = [(\alpha_1, \alpha_2, \dots, \alpha_i, \dots, \alpha_n), t],$$

где $\alpha_i = |K|_{p_i}$, $i = 1, \dots, n$; p_1, p_2, \dots, p_n — модули МСС (простые числа), такие, что $2 < p_1 < p_2 < \dots < p_n$.

Рассмотрим условия выбора модулей МСС.

2. Диапазон представления чисел

Пусть P произведение модулей МСС:

$$P = \prod_{i=1}^n p_i$$

Так как p_1, p_2, \dots, p_n — простые числа, большие двух, и их произведение является нечетным числом, то число $P - 1$ является четным и для представле-

¹ Работа выполнена при поддержке гранта президента для молодых докторов наук МД-739.2013.9.

ния положительных и отрицательных чисел в МСС используются диапазоны:

$$\left[0, \dots, \frac{1}{2}(P-1)\right] \text{ — для положительных чисел;}$$

$$\left[\frac{1}{2}(P-1)+1, \dots, P-1\right] \text{ — для отрицательных чисел.} \quad (2)$$

В МСС возможно представить все целые числа диапазона:

$$\left[-\frac{1}{2}(P-1), \dots, \frac{1}{2}(P-1)\right]. \quad (3)$$

В связи с тем, что округление является сложной операцией в МСС, оно выполняется не после каждой операции, а после группы операций, т. е. проводится отложенное округление.

Округление проводится, когда результат выходит за пределы допустимого диапазона в МСС, определяемого диапазоном (3). Следовательно, частота округлений в процессе высокоточных вычислений зависит от диапазона представления чисел. Для точного представления результата произведения двух чисел в МСС необходимо, чтобы округление выполнялось в случае его выхода не из диапазона (3), а из следующего диапазона:

$$\left[-\frac{1}{2}(P_2-1), \dots, \frac{1}{2}(P_2-1)\right], \quad (4)$$

где

$$P_2 = \prod_{i=1}^{\lfloor n/2 \rfloor} p_i$$

Рассмотрим правила выполнения арифметических операций в данном формате [6].

3. Правила выполнения арифметических операций в МСС

Так как деление в МСС является сложной операцией, то высокоточные вычисления в модулярной арифметике имеет смысл рассматривать применительно к тем задачам, в которых не используется деление вообще или деление только на константы. В последнем случае, деление заменится умножением на обратные к этим константам, вычисленные заранее.

В данной работе рассмотрены три арифметические операции: сложение, вычитание и умножение. Арифметические операции с числами вида (1) выполняются по правилам формата с плавающей точкой, но отличаются тем, что при выравнивании порядков (сложении или вычитании) мантисса большего числа сдвигается вправо, а не влево на величину, равную разности порядков большего и меньшего чисел, а порядок большего уменьшается на величину сдвига. Такой способ выравнивания порядков выбран ввиду сложности операции сдвига влево или деления в МСС.

Правила выполнения арифметических операций сложения, вычитания и умножения в модуляр-

ной системе счисления с числами A_1, A_2 и результатом A_3 , где

$$A_1 = [(\alpha_1, \alpha_2, \dots, \alpha_i, \dots, \alpha_n), t_1];$$

$$A_2 = [(\beta_1, \beta_2, \dots, \beta_i, \dots, \beta_n), t_2];$$

$$A_3 = [(\gamma_1, \gamma_2, \dots, \gamma_i, \dots, \gamma_n), t_3],$$

представлены ниже.

Сложение

Шаг № 1. Вычислить: $\gamma_i \equiv |10^{t_1 - \min(t_1, t_2)} \cdot \alpha_i + 10^{t_2 - \min(t_1, t_2)} \cdot \beta_i|_{p_i}, i = 1, \dots, n.$

Шаг № 2. $t_3 = \min(t_1, t_2).$

Вычитание

Шаг № 1. Вычислить: $\gamma_i \equiv |10^{t_1 - \min(t_1, t_2)} \cdot \alpha_i - 10^{t_2 - \min(t_1, t_2)} \cdot \beta_i|_{p_i}, i = 1, \dots, n.$

Шаг № 2. $t_3 = \min(t_1, t_2).$

Умножение

Шаг № 1. Вычислить: $\gamma_i = |\alpha_i \cdot \beta_i|_{p_i}, i = 1, \dots, n.$

Шаг № 2. $t_3 = t_1 + t_2.$

Операция округления выполняется тогда, когда результат выходит за пределы допустимого диапазона. Рассмотрим в следующем пункте способ обнаружения необходимости округления.

4. Способ обнаружения необходимости округления

Пусть

$$A_1 = K_1 \cdot 10^{t_1};$$

$$A_2 = K_2 \cdot 10^{t_2},$$

или в МСС

$$A_1 = [(\alpha_1, \alpha_2, \dots, \alpha_i, \dots, \alpha_n), t_1];$$

$$A_2 = [(\beta_1, \beta_2, \dots, \beta_i, \dots, \beta_n), t_2].$$

Тогда

$$A_1 \cdot A_2 = 10^{\lg K_1 + \lg K_2} \cdot 10^{t_1} \cdot 10^{t_2} = 10^{\lg K_1 + \lg K_2 + t_1 + t_2}$$

и справедливо неравенство

$$A_1 \cdot A_2 = K_1 \cdot 10^{t_1} + K_2 \cdot 10^{t_2} \leq 10^{\lceil \lg K_1 \rceil} \cdot 10^{t_1} + 10^{\lceil \lg K_2 \rceil} \cdot 10^{t_2} = 10^{\lceil \lg K_1 \rceil + t_1} + 10^{\lceil \lg K_2 \rceil + t_2} \leq 2 \cdot 10^{\max(\lceil \lg K_1 \rceil + t_1, \lceil \lg K_2 \rceil + t_2)} = 10^{\lg 2 + \max(\lceil \lg K_1 \rceil + t_1, \lceil \lg K_2 \rceil + t_2)}.$$

Из последнего неравенства следует, что если после выполнения арифметической операции выполняются следующие условия:

для сложения или вычитания:

$$\lg 2 \max(\lceil \lg K_1 \rceil + t_1, \lceil \lg K_2 \rceil + t_2) > \left\lceil \lg \frac{1}{2}(P_2 - 1) \right\rceil; \quad (5)$$

для умножения:

$$\lg K_1 + \lg K_2 + t_1 + t_2 > \left\lceil \lg \frac{1}{2} (P - 1) \right\rceil, \quad (6)$$

то округление результата необходимо.

Рассмотрим в следующем разделе операцию округления на основе знакоразрядной системы счисления.

5. Операция округления в МСС

Пусть при выполнении арифметических операций в МСС получен результат $A_3 = [(\gamma_1, \gamma_2, \dots, \gamma_f, \dots, \gamma_n), t_3]$, требующий округления. Округление проводится путем отбрасывания разрядов дробной части результата до n_f цифр в мантиссе A_3 и корректировке порядка t_3 .

Запишем мантиссу $K(A_3)$ числа A_3 в виде:

$$K(A_3) = \sum_{i=1}^n B_i \gamma_i - \text{rank} P, \quad (7)$$

где B_i — ортогональные базисы; rank — ранг, наибольшее положительное целое число, такое, что $K < P$.

Ортогональные базисы B_i являются константами для МСС с заданными модулями и определяются по формулам

$$\begin{aligned} B_i &= m_i \frac{P}{p_i}; \\ m_i &= \left\lfloor \frac{P}{p_i p_i} \right\rfloor^{-1}. \end{aligned} \quad (8)$$

Максимально возможное значение rank определяется из следующих неравенств:

$$\begin{aligned} \sum_{i=1}^n B_i \beta_i &\leq \sum_{i=1}^n B_i (p_i - 1) < \\ < \sum_{i=1}^n B_i p_i &= \sum_{i=1}^n m_i \frac{P}{p_i} p_i = \sum_{i=1}^n m_i P = P \left(\sum_{i=1}^n p_i - n \right). \end{aligned}$$

Отсюда видно, что

$$\text{rank} < \left(\sum_{i=1}^n p_i - n \right).$$

Рассмотрим вспомогательный алгоритм определения ранга $\text{rank} K(A_3)$ на основе позиционной характеристики, используемый при округлении.

Из выражения (7) следует, что

$$\frac{K(A_3)}{P} = \sum_{i=1}^n \frac{B_i \gamma_i}{P} - \text{rank}, \quad (9)$$

и ранг определяется следующим образом:

$$\text{rank} = \left\lfloor \frac{K(A_3)}{P} \right\rfloor = \left\lfloor \sum_{i=1}^n \frac{B_i \gamma_i}{P} \right\rfloor. \quad (10)$$

Если $K(A_3) < 0$, то из диапазона (2) следует, что значение выражения (9) больше чем 0,5 и меньше 1, если $K(A_3) > 0$, то это значение больше нуля и меньше 0,5.

Так как $K(A_3)$ принадлежит диапазону (4), то

$$\begin{aligned} \left\lfloor \frac{K(A_3)}{P} \right\rfloor_1 &\in \left[-\frac{1}{2P} (P_2 - 1), \dots, \frac{1}{2P} (P_2 - 1) \right] \subset \\ &\subset \left[\frac{\prod_{i=1}^{\lfloor n/2 \rfloor} p_i}{2 \prod_{i=1}^{\lfloor n/2 \rfloor} p_i}, \dots, \frac{\prod_{i=1}^{\lfloor n/2 \rfloor} p_i}{2 \prod_{i=1}^{\lfloor n/2 \rfloor} p_i} \right] \subset \\ &\subset \left[\frac{1}{2 \prod_{i=\lfloor n/2 \rfloor + 1}^n p_i}, \dots, \frac{1}{2 \prod_{i=\lfloor n/2 \rfloor + 1}^n p_i} \right]. \end{aligned}$$

В МСС при представлении отрицательных чисел как дополнения до модуля получим, что

$$\begin{aligned} \left\lfloor \frac{K(A_3)}{P} \right\rfloor_1 &\in \\ &\in \left[1 - \frac{1}{2 \prod_{i=\lfloor n/2 \rfloor + 1}^n p_i}, \dots, 1 + \frac{1}{2 \prod_{i=\lfloor n/2 \rfloor + 1}^n p_i} \right]. \end{aligned} \quad (11)$$

Из выражения (11) следует, например, что при $n = 10$ и значениях модулей порядка 10^5 , для отрицательных значений

$$\left\lfloor \frac{K(A_3)}{P} \right\rfloor_1 \in \left[1 - \frac{1}{2 \cdot 10^{25}}, \dots, 1 \right),$$

для положительных значений

$$\left\lfloor \frac{K(A_3)}{P} \right\rfloor_1 \in \left[1, \dots, 1 + \frac{1}{2 \cdot 10^{25}} \right).$$

Очевидно, что вследствие ошибок округления при вычислении значения выражения (10) в формате с плавающей точкой с двойной точностью значение rank для отрицательных чисел будет на единицу больше, чем истинное значение, начиная с некоторого n . Величины B_i/P являются константами и могут быть вычислены в формате с плавающей точкой заранее.

Схема для быстрого определения ранга представлена на рис. 1. В соответствии с этой схемой значение $\text{rank} K(A_3)$ может быть вычислено по формуле (10) за $\lceil \log_2 n \rceil + 1$ шагов.

После определения ранга $\text{rank} K(A_3)$ для округления результата A_3 , как было указано выше, требуется отбросить все цифры дробной части результата до n_f цифр в мантиссе A_3 и скорректировать порядок t_3 .

Округление мантиссы $K(A_3)$ числа A_3 , определяемой по формуле (7), требует вычислений с плавающей точкой с точностью до $\lceil \lg P \rceil$ цифр, например, для $n = 10$ и модулей порядка 10^5 точность вычислений — 50 десятичных цифр. Очевидно, что такая точность вычислений резко увеличит общее время операции округления.

Рассмотрим знакоразрядную систему счисления с цифрами в диапазоне $[-6, \dots, 6]$, в которой возможно ускоренное вычисление значения выражения (7).

Сложение чисел в этой системе счисления $(x_{n-1}, \dots, x_0) + (y_{n-1}, \dots, y_0) = (s_{n-1}, \dots, s_0)$ про-

водится справа налево и осуществляется в два этапа по формулам

$$u_i = x_i + y_i - 10c_i,$$

где

$$c_i = \begin{cases} 1, & \text{если } (x_i + y_i \geq 6, \\ -1, & \text{если } (x_i + y_i \leq -6, \\ 0, & \text{если } |x_i + y_i| < 6, \end{cases}$$

и результат суммы определяется по формуле

$$s_i = u_i + c_{i-1}.$$

В знакоразрядной системе счисления исключается распространение переноса от старшего разряда к младшему, благодаря чему арифметические операции сложения, вычитания распараллеливаются и время их выполнения не зависит от разрядности входных данных.

Рассмотрим способ ускоренного вычисления значения выражения (7) в знакоразрядной системе счисления с цифрами в диапазоне $[-6, \dots, 6]$.

Пусть константы B_j, P заданы в знакоразрядной системе счисления.

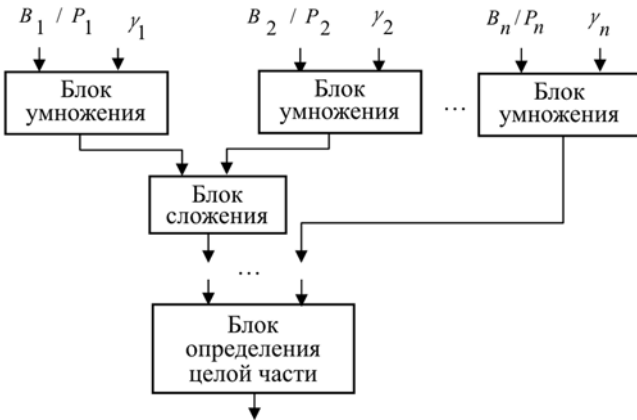


Рис. 1. Схема для быстрого вычисления ранга

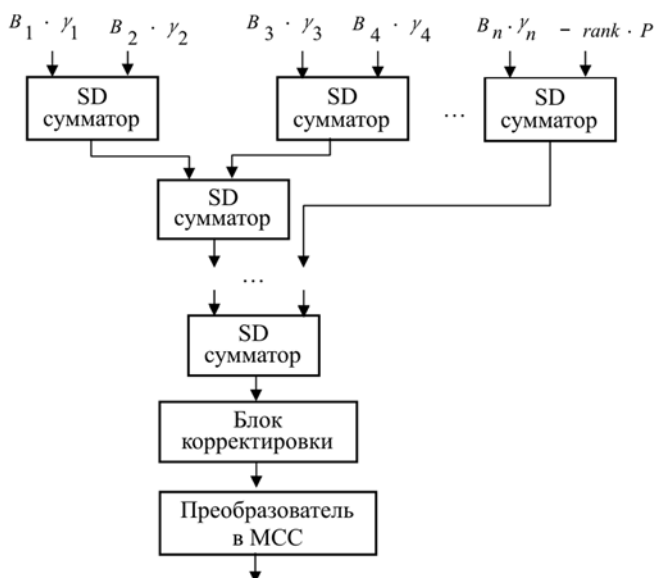


Рис. 2. Схема параллельного округления

Выражения вида

$$jB_j, j = 1 \dots p_i - 1;$$

$$jP, j = 1 \dots \left(\sum_{i=1}^n p_i - n \right) \quad (12)$$

являются константами и могут храниться в памяти ЭВМ, тогда процесс округления можно представить в виде схемы, приведенной на рис. 2.

В соответствии с приведенной схемой $K(A_3)$ может быть вычислено по формуле (7) за $\lceil \log_2 n \rceil + 1$ шагов при известном значении $rank$. Блок корректировки, представленный на рис. 2, выполняет отбрасывание старших разрядов результата до n_f цифр и уменьшает порядок t_3 на число отброшенных цифр. Последний блок преобразует полученный после округления результат в МСС.

Общая оценка числа шагов для округления результата в соответствии со схемой рис. 2 оценивается следующим выражением:

$$\lceil \log_2 n \rceil + 2(\lceil \log_2 n \rceil + 1) + 1 + \lceil \log_2 n \rceil = 4\lceil \log_2 n \rceil + 3.$$

Недостатком приведенного способа округления является необходимость хранения в памяти ЭВМ массивов констант (12) по каждому модулю. Рассмотрим способ уменьшения числа хранимых констант.

Запишем γ_i и $rank$ в десятичной системе счисления в следующем виде:

$$\gamma_i = \gamma_i^0 \cdot 10^{n_1} + \gamma_i^1 \cdot 10^{n_1-1} + \dots + \gamma_i^{n_1}, i = 1 \dots n, \quad (13)$$

$$rank = r_0 \cdot 10^{n_2} + r_1 \cdot 10^{n_2-1} + \dots + r_{n_2},$$

где n_1, n_2 — целые числа, такие, что

$$n_1 = \lfloor \lg p_i \rfloor, n_2 = \left\lceil \lg \left(\sum_{i=1}^n p_i - n \right) \right\rceil,$$

$$0 \leq \gamma_i^j \leq 9, i = 1 \dots n, j = 1 \dots n_1, 0 \leq r_j \leq 9, j = 1 \dots n_2.$$

Подставим (13) в выражение (7) и получим

$$K(A_3) = \sum_{i=1}^n B_i (\gamma_i^0 \cdot 10^{n_1} + \gamma_i^1 \cdot 10^{n_1-1} + \dots + \gamma_i^{n_1}) - (r_0 \cdot 10^{n_2} + r_1 \cdot 10^{n_2-1} + \dots + r_{n_2})P, \quad (14)$$

$$j = 1 \dots 9.$$

При нахождении значения $K(A_3)$ в формуле (14) по схеме, представленной на рис. 2, требуется хранение меньшего числа констант, и суммирование выполняется параллельно по степеням десятки.

Заключение

Как было указано выше, округление в модулярной системе счисления является сложной операцией. Предложен способ округления в комплексе систем счисления МСС и знакоразрядной, позволяющий распараллелить операцию округления. Данный способ применим при реализации других немодульных операций. Наибольший эффект ускорения достигается при аппаратной реализации параллельного округления.

Список литературы

1. **Bailey D. H.** High-Precision Computation and Mathematical Physics // Lawrence Berkeley National Laboratory, 2009. URL: <http://crd-legacy.lbl.gov/~dhbailey/dhbpapers/dhb-jmb-acat08.pdf>
2. **Соловьев Р. А., Балака Е. С., Тельпухов Д. В.** Устройство для вычисления скалярного произведения векторов с коррекцией ошибок на базе системы остаточных классов // Проблемы разработки перспективных микро- и наноэлектронных систем — 2014. Сборник трудов / Под общ. ред. А. Л. Стемпковского. М.: Изд. ИППМ РАН, 2014. Часть IV. С. 173—178.
3. **Phatak D. S., Koren I.** Hybrid Signed Digit Number Systems: A Unified Framework for Redundant Number Representations with

Bounded Carry Propagation Chains // IEEE Transactions on Computers. August 1994. V. 43. N. 8. P. 880—891.

4. **Дзегеленок И. И., Оцоков Ш. А.** Алгебраизация числовых представлений в обеспечении высокоточных суперкомпьютерных вычислений // Вестник МЭИ. 2010. № 3. С. 107—116.
5. **Оцоков Ш. А.** Эффективный алгоритм округления в высокоточных вычислениях в модулярной арифметике // Информационные технологии. 2013. № 10. С. 35—39.
6. **Исупов К. С.** Модулярно-позиционный формат и программный пакет для разрядно-параллельных вычислений высокой точности // Вестник Южноуральского государственного университета. 2013. № 1.

Sh. A. Otsokov. Associate Professor, e-mail: shamil24@mail.ru,
National Research University "MPEI"

Acceleration of High-Precision Computation Based on Parallelization of Group Number Systems

In article the method of acceleration of high-precision computation based on parallelization of two numeral systems is offered: modular and signed-digital. Advantages of this approach are connected with high speed of not modular operations, such as the back transformation to a position numeral system from residue number system, comparisons of numbers, definition of a sign.

Keywords: modular arithmetic, high accuracy computation, redundant number system

References

1. **Bailey D. H.** High-Precision Computation and Mathematical Physics. Lawrence Berkeley National Laboratory, 2009. URL: <http://crd-legacy.lbl.gov/~dhbailey/dhbpapers/dhb-jmb-acat08.pdf>
2. **Solov'ev R. A., Balaka E. S., Tel'pukhov D. V.** Ustroistvo dlya vychisleniya skalyarnogo proizvedeniya vektorov s korrktsiei oshibok na baze sistemy ostatochnykh klassov. *Problemy razranotki perspektivnykh mikro- i nanoelektronnykh sistem—2014. Sbornik trudov pod obshch. red. ak. RAN A. L. Stempkovskogo. M.: IPPM RAN, 2014. Chast' IV. P. 173—178.*
3. **Phatak D. S., Koren I.** Hybrid Signed Digit Number Systems: A Unified Framework for Redundant Number Representations with

Bounded Carry Propagation Chains. *IEEE Transactions on Computers.* August 1994. V. 43. N. 8. P. 880—891.

4. **Dzegelenok I. I., Otsokov Sh. A.** Algebraizatsiya chislovykh predstavlenii v obespechenii vysokotochnykh superkomp'yuternykh vychislenii. *Vestnik MEI.* 2010. N. 3. P. 107—116.
5. **Otsokov Sh. A.** Effektivnyi algoritm okrugleniya v vysokotochnykh vychisleniyakh s modulyarnoi arifmetike. *Informatsionnye tekhnologii.* 2013. N. 10. P. 35—39.
6. **Isupov K. S.** Modulyarno-pozitsionnyi format i programmnyi paket dlya razryadno-parallelnykh vychislenii vysokoi tochnosti. *Vestnik Yuzhnoural'skogo gosudarstvennogo universiteta.* 2013. N. 1. P. 65—80.

УДК 004.272.42

Р. Н. Федюнин, канд. техн. наук, доц., e-mail: frn_penza@mail.ru
Пензенский государственный университет, г. Пенза

Оценка пространственной сложности функциональных блоков АЛУ на базе однородных вычислительных структур

Рассмотрен способ оценки пространственной сложности вычислительных модулей на базе систолических структур. Данный подход позволяет быстро и адекватно провести оценку аппаратных затрат на реализацию вычислительных алгоритмов в рамках однородных вычислительных структур как позиционной, так и модулярной арифметики.

Ключевые слова: пространственная сложность, систолическая структура, модулярная позиционная арифметика, класс вычислений

При проектировании вычислительных устройств разработчики, помимо прочих, выделяют две задачи: расчет временной и пространственной (в данном примере она же — аппаратная) сложности вычислений [1]. В данной работе рассмотрен способ оценки

пространственной сложности вычислительных модулей [2] на базе систолических структур [3—5]. Данный подход позволяет быстро и адекватно провести теоретическую оценку аппаратных затрат на реализацию вычислительных алгоритмов в рамках

однородных вычислительных структур (ОВС) как позиционной [3], так и модулярной арифметики [4].

Для иллюстрации описываемого способа оценки пространственной сложности в качестве примера возьмем схему VLIW-процессора (рис. 1).

Каждый блок АЛУ процессора — однородная вычислительная структура класса А [5] (рис. 2), пространственную сложность которой требуется оценить для последующего сравнительного анализа с аналогами.

К классу А вычислений относятся вычисления, для которых справедливы приведенные ниже положения.

1. Обрабатываемая структура данных состоит из N элементов, каждый из которых представлен M -разрядным двоичным кодом.

2. На каждом шаге вычислений выполняемая P_n операция преобразования элемента данных из набора P является итерационно-разрядной операцией: вычисляемое значение i -го разряда зависит от вычисленного значения $(i - 1)$ -го разряда.

3. При выполнении вычислений P обработка элементов структуры N — массовая: одна и та же операция преобразования P_n применяется ко всем N элементам обрабатываемой структуры данных.

4. При выполнении вычислений P обработка данных N -элементной структуры — итерационно-независимая: значение результата обработки i -го элемента данных не зависит от результата обработки $(i - 1)$ -го элемента данных, порядок поступления данных на обработку не имеет значения.

Отметим, что в работах [6—8] рассмотрен расчет пространственной сложности только для вычислительных структур в позиционной системе счисления, а рассматриваемый в данной работе способ справедлив как для позиционной, так и для модулярной обработки данных.

Отличия топологий ОВС в позиционной системе счисления и модулярной ОВС в том, что ОВС в позиционной системе счисления обрабатывает число определенной разрядности, а значит ее размерность по ширине и высоте определяется разрядностью данных. Рис. 3 иллюстрирует прямо пропорциональное увеличение аппаратных затрат при увеличении разрядности ОВС в позиционной системе счисления — увеличение разрядности данных в 2 раза требует увеличения аппаратных ресурсов ОВС примерно в 4 раза.

В свою очередь, для обработки тех же данных, но в системе остаточных классов (СОК), обработка ведется не самих исходных данных, а остатков от данных по конкретным модулям [8]. Так, в работах [10—12] показано, что для обработки 32-разрядного позиционного числа требуются 7×7 разрядные модулярные ОВС и 7-разрядные регистры хранения данных и результата. При этом число требуемых для этого параллельно работающих блоков увеличивается не в несколько раз, а на константу, значение которой определяется разрядностью данных [8, 9]. С учетом рассмотренной специфики графиче-

ское представление матрицы ОВС в СОК иллюстрируется рис. 4.

Отметим, что последующий анализ не учитывает временные и аппаратные затраты на преобразование данных из позиционной системы счисления в СОК и обратно, данное исследование подробно проведено в работах [3, 12].

Фактором оценки аппаратных затрат на реализацию ОВС в различных системах счисления является число ячеек, которое требуется для реализации того или иного типа вычислений.

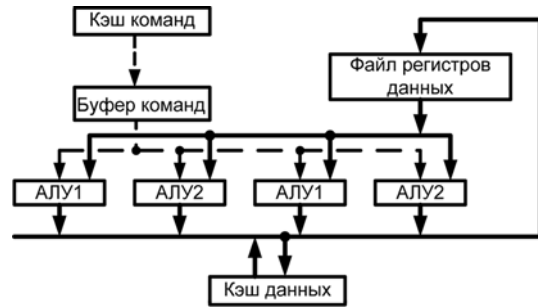


Рис. 1. Обобщенная структура операционной части VLIW-процессора

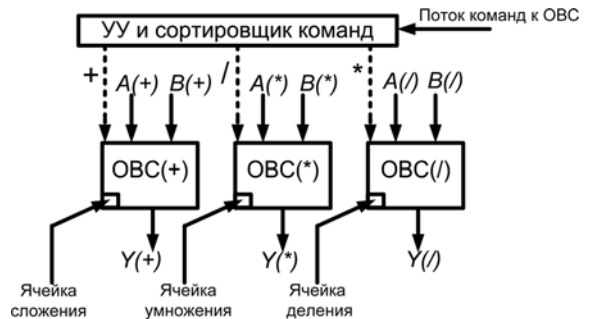


Рис. 2. Вариант АЛУ на базе ОВС

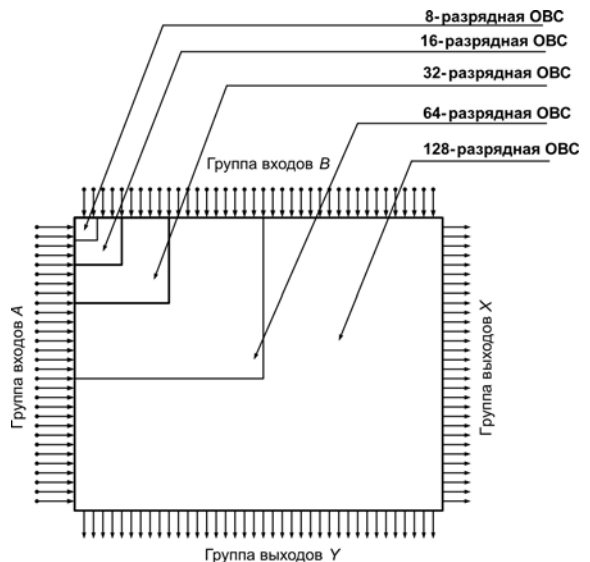


Рис. 3. Увеличение аппаратных затрат при увеличении разрядности ОВС в позиционной системе счисления



Рис. 4. Увеличение аппаратных затрат при увеличении разрядности ОВС в СОК

Предположим, что одна ячейка ОВС — это единица аппаратных затрат на реализацию ОВС при соблюдении следующих условий [6].

1. Операционные части исследуемых ОВС (ОЧОВС) выполняют однотипные операции.

2. Настройка операционной среды выполняется устройством управления ОВС, а ход выполнения операции в ОЧОВС определяется промежуточными значениями данных, обрабатываемых в ОЧОВС.

При выполнении данных условий можно, подсчитав число ячеек, требуемых для выполнения ОЧОВС той или иной операции, рассчитать ее аппаратные затраты без учета аппаратных затрат на реализацию устройства управления.

Для количественной оценки аппаратных затрат на реализацию операционной части ОВС в позиционной системе счисления и СОК отметим особенности организации ОВС в данных системах счисления.

Как показано в работах [13–15] ОВС для реализации арифметических операций в позиционной системе счисления представляет собой матрицу $N \times N$, где N — число столбцов и строк операционной части ОВС, необходимое для реализации

требуемой операции, и разрядность обрабатываемых данных.

Операционная часть ОВС в СОК представляет собой матрицу $N \times M$, где N — число строк матрицы и разрядность данных, а M — число столбцов матрицы операционной части ОВС в СОК, где $M \geq N$, это связано со спецификой выполнения операций в модулярной арифметике [12]. Так, после выполнения операции в СОК возможна коррекция результата, максимальное число тактов коррекции — два [12]. Таким образом, на каждый столбец матрицы в позиционной ОВС приходится три столбца в ОВС СОК. Если рассматривать число строк ОЧОВС, которое в обоих случаях равно разрядности данных, то получим, что ОВС в СОК обрабатывают данные разрядностью не более 7 бит [12] при условии, что исходные данные имели разрядность не более 128 разрядов. При этом в работе [13] доказано и показано на примере, что возможно совмещение в одном столбце как обработки данных, так и коррекции результата, таким образом полноценная обработка данных при корректном размещении связей внутри столбца конвейера и между его столбцами дает нам $M = N$. Если разрядность исходных данных превышает 128-разрядный диапазон, то для их обработки требуются ОЧОВС для СОК 8- и более разрядные [12, 13], получается или одна большая матрица в позиционной системе счисления (см. рис. 3), или набор небольших матриц, каждая из которых работает с данными по конкретному модулю (см. рис. 4). Аппаратные затраты на реализацию операционной части ОВС в позиционной системе счисления и ОВС в СОК рассчитываются из следующих соображений [13] (табл. 1).

Если матрица ОЧОВС реализации в позиционной системе счисления, то формула расчета ее трудоемкости равна $N \cdot N$.

Для матриц ОЧОВС реализации в СОК выражение для расчета аппаратной сложности зависит от разрядности исходных данных. Так, для реализации ОЧОВС в СОК для работы с 8-разрядными данными требуется одна матрица ОЧОВС по модулю 2 и одна матрица ОЧОВС по модулю 7. Число и разрядность ОЧОВС для работы с данными той или иной разрядности рассчитываются путем деления максимально возможного числа заданной разрядности на модули СОК, начиная с самого боль-

Таблица 1

Объем аппаратных затрат при реализации вычислений в ОЧОВС

Разрядность данных, поступивших на обработку	Размерность ОВС в позиционной системе счисления, необходимая для обработки данных	Число ячеек ОВС в позиционной системе счисления, необходимое для обработки данных	Разрядность данных, поступивших на обработку	Размерность ОВС в СОК, необходимая для обработки данных	Число ячеек ОВС в СОК, необходимое для обработки данных
8	8×8	64	1 + 7	1 × 3 и 7 × 7	50
16	16×16	256	2 + 7	2 × 6 и 7 × 7	65
32	32×32	1024	$4 \cdot 7 + 6 + 2$	7×7 и 6×6 и 1×3	235
64	64×64	4096	$1 + 2 \cdot 6 + 8 \cdot 7$	1×3 и 6×6 и 7×7	467
128	128×128	16384	$8 \cdot 7 + 7 \cdot 6 + 7 \cdot 5 + 2 \cdot 4 + 2 \cdot 3 + 1$	7×7 и 6×6 и 5×5 и 4×4 и 3×3 и 1×3	872

шого модуля, на который возможно поделить требуемое число. Итерационный процесс деления на рабочие модули СОК прекращается при условии, что полученный остаток меньше двух, так как два — минимальный рабочий модуль системы остаточных классов [1]. Данная методика подсчета числа матриц ОЧОВС, требуемых для организации модульных вычислений, и их размерности применима для данных с любой разрядностью. Отметим, что для выполнения арифметических операций по заданному модулю возможно применение нескольких матриц одной размерности. Так, для организации 32-разрядных вычислений требуются четыре 7-разрядные матрицы, каждая из которых настроена на работу с индивидуальным модулем (см. табл. 1). Расчеты, приведенные в табл. 1 и на рис. 5, показывают, что аппаратные затраты позиционных ОВС в несколько раз больше аппаратных затрат ОВС СОК.

Расчеты пространственной сложности ОВС в позиционной системе счисления на основе формул и свойств вычисления, проводят по следующим формулам:

$$\sum_{i=1}^n K \cdot SPACE(M_i \cdot N_i + 2D_i); \quad (1)$$

$$\sum_{i=1}^n K \cdot SPACE(M_i \cdot N_i + D_i(D_i - 1)); \quad (2)$$

$$\sum_{i=1}^n K \cdot SPACE(M_i \cdot N_i \cdot D_i + 2D_i); \quad (3)$$

где индексы элементов M , D , N определяют разрядность соответствующих элементов формул, при этом K — это количество однородных вычислительных структур с параметрами M_i , N_i и D_i . Формулы (1)—(3) учитывают специфику представления данных, так (1) — формула пространственной сложности ОВС при массово-последовательных разрядно-параллельных вычислениях [2] и при массово-последовательных разрядно-последовательных вычислениях; (2) — формула пространственной сложности ОВС при массово-параллельных разрядно-последовательных вычислениях [2]; (3) — формула пространственной сложности ОВС при массово-параллельных разрядно-параллельных вычислениях [2, 3].

На основании формул (1)—(3) требуется получить формулы, учитывающие специфику построения операционной части ОВС в СОК, так как в отличие от позиционной системы счисления ОВС в СОК состоит зачастую из нескольких вычислительных матриц.

В табл. 2—5 представлены правила формирования вычислительного пространства ОВС СОК для различных типов вычислений.

По формулам, приведенным в табл. 2—5, рассчитаем численные значения пространственной сложности матричных вычислений и сведем их в единой табл. 6.

Расчеты, приведенные в табл. 6, показывают, что для работы с N -разрядными данными в позиционной системе счисления требуются ОВС, аппа-

ратные затраты которых многократно возрастают с увеличением разрядности данных. В то время как в СОК, используя матрицу по максимальному модулю, с каналами подстройки среды для работы по конкретному модулю возможно существенное сокращение аппаратных затрат, требуемых для обработки информации (для работы с числами вплоть до 256 разрядов — это 7-разрядный модуль [1]). Наглядная иллюстрация данного вывода представлена на рис. 5.

Таблица 2

Массово-последовательные разрядно-параллельные вычисления в ОВС СОК

№	Разрядность обрабатываемых данных, бит	Число (шт.) и разрядность ОВС СОК, бит	Формулы для расчета пространственной сложности
1	8	1 ОВС — 7 × 7 1 ОВС — 1 × 1	$SPACE(M1 \cdot N1 + 2 \cdot D1) + SPACE(M7 \cdot N7 + 2 \cdot D7)$
2	16	1 ОВС — 7 × 7, 1 ОВС — 2 × 2	$SPACE(M2 \cdot N2 + 2 \cdot D2) + SPACE(M7 \cdot N7 + 2 \cdot D7)$
3	32	4 ОВС — 7 × 7, 1 ОВС — 6 × 6, 1 ОВС — 2 × 2	$4 \cdot SPACE(M7 \cdot N7 + 2 \cdot D7) + SPACE(M6 \cdot N6 + 2 \cdot D6) + SPACE(M2 \cdot N2 + 2 \cdot D2)$
4	64	8 ОВС — 7 × 7, 2 ОВС — 6 × 6, 1 ОВС — 1 × 1	$SPACE(M1 \cdot N1 + 2 \cdot D1) + 2 \cdot SPACE(M6 \cdot N6 + 2 \cdot D6) + 8 \cdot SPACE(M7 \cdot N7 + 2 \cdot D7)$
5	128	8 ОВС — 7 × 7, 7 ОВС — 6 × 6, 7 ОВС — 5 × 5, 2 ОВС — 4 × 4, 1 ОВС — 3 × 3, 1 ОВС — 1 × 1	$SPACE(M1 \cdot N1 + 2 \cdot D1) + 2 \cdot SPACE(M3 \cdot N3 + 2 \cdot D3) + 2 \cdot SPACE(M4 \cdot N4 + 2 \cdot D4) + 7 \cdot SPACE(M5 \cdot N5 + 2 \cdot D5) + 7 \cdot SPACE(M6 \cdot N6 + 2 \cdot D6) + 8 \cdot SPACE(M7 \cdot N7 + 2 \cdot D7)$
Примечание. M — число строк в ОВС СОК; N — число столбцов ОВС СОК; D — разрядность обрабатываемых данных.			

Таблица 3

Массово-последовательные разрядно-последовательные вычисления в ОВС СОК

№	Разрядность обрабатываемых данных, бит	Число (шт.) и разрядность ОВС СОК, бит	Формулы для расчета пространственной сложности
1	8	1 ОВС — 7 × 7, 1 ОВС — 1 × 1	$SPACE(M1 \cdot N1 + 2 \cdot D1) + SPACE(M7 \cdot N7 + 2 \cdot D7)$
2	16	1 ОВС — 7 × 7, 1 ОВС — 2 × 2	$SPACE(M2 \cdot N2 + 2 \cdot D2) + SPACE(M7 \cdot N7 + 2 \cdot D7)$
3	32	4 ОВС — 7 × 7, 1 ОВС — 6 × 6, 1 ОВС — 2 × 2	$4 \cdot SPACE(M7 \cdot N7 + 2 \cdot D7) + SPACE(M6 \cdot N6 + 2 \cdot D6) + SPACE(M2 \cdot N2 + 2 \cdot D2)$
4	64	8 ОВС — 7 × 7, 2 ОВС — 6 × 6, 1 ОВС — 1 × 1	$SPACE(M1 \cdot N1 + 2 \cdot D1) + 2 \cdot SPACE(M6 \cdot N6 + 2 \cdot D6) + 8 \cdot SPACE(M7 \cdot N7 + 2 \cdot D7)$
5	128	8 ОВС — 7 × 7, 7 ОВС — 6 × 6, 7 ОВС — 5 × 5, 2 ОВС — 4 × 4, 2 ОВС — 3 × 3, 1 ОВС — 1 × 1	$SPACE(M1 \cdot N1 + 2 \cdot D1) + 2 \cdot SPACE(M3 \cdot N3 + 2 \cdot D3) + 2 \cdot SPACE(M4 \cdot N4 + 2 \cdot D4) + 7 \cdot SPACE(M5 \cdot N5 + 2 \cdot D5) + 7 \cdot SPACE(M6 \cdot N6 + 2 \cdot D6) + 8 \cdot SPACE(M7 \cdot N7 + 2 \cdot D7)$
Примечание. M — число строк в ОВС СОК; N — число столбцов ОВС СОК; D — разрядность обрабатываемых данных.			

Таблица 4

Массово-параллельные разрядно-последовательные вычисления в ОВС СОК

№	Разрядность обрабатываемых данных, бит	Число (шт.) и разрядность ОВС СОК, бит	Формулы для расчета пространственной сложности
1	8	1 ОВС — 7 × 7, 1 ОВС — 1 × 1	$SPACE(M1 \cdot N1 + D1(D1 - 1)) +$ $+ SPACE(M7 \cdot N7 + D7(D7 - 1))$
2	16	1 ОВС — 7 × 7, 1 ОВС — 2 × 2	$SPACE(M2 \cdot N2 + D2(D2 - 1)) +$ $+ SPACE(M7 \cdot N7 + D7(D7 - 1))$
3	32	4 ОВС — 7 × 7, 1 ОВС — 6 × 6, 1 ОВС — 2 × 2	$4 \cdot SPACE(M7 \cdot N7 + 2 \cdot D7) +$ $+ SPACE(M6 \cdot N6 + 2 \cdot D6) +$ $+ SPACE(M2 \cdot N2 + 2 \cdot D2)$
4	64	8 ОВС — 7 × 7, 2 ОВС — 6 × 6, 1 ОВС — 1 × 1	$SPACE(M1 \cdot N1 + 2 \cdot D1) +$ $+ 2 \cdot SPACE(M6 \cdot N6 + 2 \cdot D6) +$ $+ 8 \cdot SPACE(M7 \cdot N7 + 2 \cdot D7)$
5	128	8 ОВС — 7 × 7, 7 ОВС — 6 × 6, 7 ОВС — 5 × 5, 2 ОВС — 4 × 4, 2 ОВС — 3 × 3, 1 ОВС — 1 × 1	$SPACE(M1 \cdot N1 + 2 \cdot D1) +$ $+ 2 \cdot SPACE(M3 \cdot N3 + 2 \cdot D3) +$ $+ 2 \cdot SPACE(M4 \cdot N4 + 2 \cdot D4) +$ $+ 7 \cdot SPACE(M5 \cdot N5 + 2 \cdot D5) +$ $+ 7 \cdot SPACE(M6 \cdot N6 + 2 \cdot D6) +$ $+ 8 \cdot SPACE(M7 \cdot N7 + 2 \cdot D7)$

Примечание. *M* — число строк в ОВС СОК; *N* — число столбцов ОВС СОК; *D* — разрядность обрабатываемых данных

Таблица 5

Массово-параллельные разрядно-параллельные вычисления в ОВС СОК

№	Разрядность обрабатываемых данных, бит	Число (шт.) и разрядность ОВС СОК, бит	Формулы для расчета пространственной сложности
1	8	1 ОВС — 7 × 7, 1 ОВС — 1 × 1	$SPACE(M1 \cdot N1 \cdot D1 + 2 \cdot D1) +$ $+ SPACE(M7 \cdot N7 \cdot D7 + 2 \cdot D7)$
2	16	1 ОВС — 7 × 7, 1 ОВС — 2 × 2	$SPACE(M2 \cdot N2 \cdot D2 + 2 \cdot D2) +$ $+ SPACE(M7 \cdot N7 \cdot D7 + 2 \cdot D7)$
3	32	4 ОВС — 7 × 7, 1 ОВС — 6 × 6, 1 ОВС — 2 × 2	$4 \cdot SPACE(M7 \cdot N7 \cdot D7 + 2 \cdot D7) +$ $+ SPACE(M6 \cdot N6 \cdot D6 + 2 \cdot D6) +$ $+ SPACE(M2 \cdot N2 \cdot D2 + 2 \cdot D2)$
4	64	8 ОВС — 7 × 7, 2 ОВС — 6 × 6, 1 ОВС — 1 × 1	$SPACE(M1 \cdot N1 + 2 \cdot D1) +$ $+ 2 \cdot SPACE(M6 \cdot N6 + 2 \cdot D6) +$ $+ 8 \cdot SPACE(M7 \cdot N7 + 2 \cdot D7)$
5	128	8 ОВС — 7 × 7, 7 ОВС — 6 × 6, 7 ОВС — 5 × 5, 2 ОВС — 4 × 4, 2 ОВС — 3 × 3, 1 ОВС — 1 × 1	$SPACE(M1 \cdot N1 + 2 \cdot D1) +$ $+ 2 \cdot SPACE(M3 \cdot N3 + 2 \cdot D3) +$ $+ 2 \cdot SPACE(M4 \cdot N4 + 2 \cdot D4) +$ $+ 7 \cdot SPACE(M5 \cdot N5 + 2 \cdot D5) +$ $+ 7 \cdot SPACE(M6 \cdot N6 + 2 \cdot D6) +$ $+ 8 \cdot SPACE(M7 \cdot N7 + 2 \cdot D7)$

Примечание. *M* — число строк в ОВС СОК; *N* — число столбцов ОВС СОК; *D* — разрядность обрабатываемых данных.

В конечном итоге, в результате исследования получен способ теоретической оценки пространственной сложности, адекватность которого была проверена реализацией вычислительных устройств на базе различных типов ПЛИС, с последующим анализом аппаратных затрат в САПР Aldec Active HDL. Практичность способа доказана неоднократным использованием в рамках грантов.

Таблица 6

Расчеты пространственной сложности ОВС

Способ представления данных	Разрядность данных	Численные значения пространственной сложности позиционной ОВС	Разрядность данных	Численные значения пространственной сложности ОВС в СОК
Массово-последовательные разрядно-параллельные вычисления	8	80	9	66
	16	288	10	71
	32	1088	14	308
	64	4224	14	608
	128	16640	17	1118
Массово-последовательные разрядно-последовательные вычисления	8	80	9	66
	16	288	10	71
	32	1088	14	308
	64	4224	14	608
	128	16640	17	1118
Массово-последовательные разрядно-последовательные вычисления	8	120	9	92
	16	496	10	97
	32	2016	14	434
	64	8128	14	861
	128	32640	17	1607
Массово-последовательные разрядно-последовательные конвейерные вычисления	8	528	9	360
	16	4128	10	369
	32	3283	14	1668
	64	26230	14	3315
	128	209700	17	5610

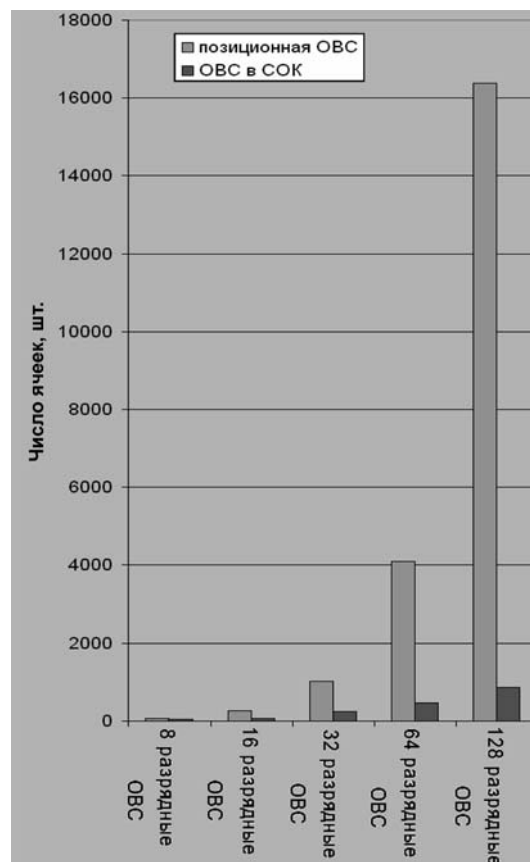


Рис. 5. Объем аппаратных затрат при реализации вычислений в ОЧОВС

Несмотря на наличие в средствах реализации СБИС систем программных модулей для расчета и визуализации аппаратных затрат, данный способ на ранних этапах формирования архитектуры позволяет оценить адекватность аппаратных затрат и внести нужные разработчикам коррективы, также автор неоднократно применял данный способ при обосновании патентных решений на базе однородных вычислительных систем [3].

Список литературы

1. Акушский И. Я., Юдицкий Д. И. Машинная арифметика в остаточных классах. М.: Сов. радио, 1968. 439 с.
2. Федюнин Р. Н. Оценка пространственно-временной сложности и способы повышения скорости двоичных арифметических операций // Научное обозрение. 2006. № 3. С. 100—111.
3. Федюнин Р. Н., Князьков В. С. Ячейка однородной вычислительной среды / Патент РФ 2284568. Оpubл. 27.09.2006. Бюл. № 27.
4. Пат. 2491612 Российская Федерация, МПК G06F7/50. Способ организации вычислений суммы n m -разрядных чисел / Осинин И. П., Князьков В. С. Оpubл. 27.08.2013 г., Бюл. № 24.
5. Амербаев В. М., Балака Е. С., Константинов А. В., Тельпухов Д. В. Реализация обратного преобразователя модулярной арифметики, совмещенного с операцией округления для задач ЦОС // V Всероссийская научно-техническая конференция "Проблемы разработки перспективных микро- и наноэлектрон-

ных систем — 2012". Сб. науч. тр. / Под общ. ред. А. Л. Стемповского. М.: ИППМ РАН, 2012. С. 535—538.

6. Головкин Б. А. Параллельные вычислительные системы. М.: Наука; Главная редакция физмат литературы, 1980. 517 с.
7. Pramod Kumar Meher. Hardware-Efficient Systolization of DA-based Calculation of Finite Digital, Convolution Circuits and Systems II: Express Briefs, IEEE Transactions. Is. 8. Date Aug. 2006. P. 707—711.
8. Maltsev A., Pestretsov V., Maslennikov R., Khoryaev A. Triangular systolic array with reduced latency for QR-decomposition of complex matrices // Circuits and Systems, 2006. ISCAS 2006. Proceedings IEEE International Symposium. 2006. P. 602—608.
9. Карцев М. А., Брик В. А. Вычислительные системы и синхронная арифметика. М.: Радио и связь, 1981. 359 с.
10. Карцев В. А. Арифметика цифровых машин. М.: Наука, 1979. 575 с.
11. Останиевич А. В. Экспериментальное исследование поддержки предметных вычислений в архитектуре с явно выраженным параллелизмом // Информационные технологии и вычислительные системы. 1999. № 1.
12. Костин А. Е., Шангин В. Ф. Организация и обработка структур данных в вычислительных системах. М.: Высшая школа, 1987. 245 с.
13. Hegde G., Prasanna C., Raj P., Vaya P. R. Implementation of Systolic Array Architecture for Full Search Block Matching Algorithm on FPGA / European Journal of Scientific Research. 2009. Vol. 33, N. 4. P. 606—616.
14. Князьков В. С., Бикташев Р. А. Параллельные вычислительные системы с ОКМД-архитектурой: Уч. пособие. Пенза: RIO Пензенского политехнического ин-та, 1991. 51 с.
15. Федюнин Р. Н. Способ реализации аппаратного слоя VLIW-архитектуры на базе систолических структур // Известия высших учебных заведений, Поволжский регион (Технические науки). Пенза, 2013. № 2. С. 15—22.

R. N. Fedjunin, Ph. D., e-mail: frn_penza@mail.ru, Penza State University

The Evaluation of Hardware Complexity Functional Part of ALU Based Systolic Structure

When designing computing devices, developers have identified two problems: the calculation of the time and hardware complexity of computation. The article describes the method of estimating the hardware complexity of computational modules based on systolic structures.

This approach allows quickly and adequately makes the hardware costs of the computational algorithms for positioning and modular systolic structures.

Note that in studying the works of hardware complexity calculation considered only for computational structures in a positional number system, and considered in the article is just a way for both positional and for modular data processing.

For example the described method of estimating the spatial complexity, as an example, takes the VLIW processor circuit.

Each unit ALU processor is systolic structure of class A, hardware complexity is to be evaluated for subsequent comparative analysis with peers.

Keywords: hardware complexity, systolic structure, residue number system, classes of computations, very long instruction word, explicitly parallel instruction computing

References

1. Akushskij I. Ja., Judickij D. I. Mashinnaja arifmetika v ostatochnyh klassah. Moscow: Sov. Radio, 1968. 439 p.
2. Fedjunin R. N. Ocenka prostranstvenno-vremennoj slozhnosti i sposoby povysheniya skorosti dvoichnyh arifmeticheskikh operacij. Nauchnoe obozrenie. 2006. N. 3. P. 100—111.
3. Fedjunin R. N., Knjaz'kov V. S. Jachejka odnorodnoj vychislitel'noj sredy. Patent RF 2284568. Publ. 27.09.2006. Vol. N. 27.
4. Osinin I. P., Knjaz'kov V. S. Patent 2491612 Rossijskaja Federacija. MPK G06F7/50. Sposob organizacii vychislenij summy n m-razrdnyh chisel. Publ. 27.08.2013. Vol. N. 24.
5. Amerbaev V. M., Balaka E. S., Konstantinov A. V., Tel'puhov D. V. Realizacija obratnogo preobrazovatelja moduljarnoj arifmetiki sovmeshchennogo s operaciej okruglenija dlja zadach COS. V Vserossijskaja nauchno-tehnicheskaja konferencija "Problemy razrabotki perspektivnyh mikro- i nanojelektronnyh sistem — 2012". Sb. nauch. tr. / Ed. A. L. Stempkovskij. M.: IPPM RAN. 2012. P. 535—538.
6. Golovkin B. A. Parallel'nye vychislitel'nye sistemy. M.: Nayka; Glavnaia redakcija fizmat literatury, 1980. 517 p.
7. Pramod Kumar Meher. Hardware-Efficient Systolization of DA-based Calculation of Finite Digital. Convolution Circuits and Systems II: Express Briefs, IEEE Transactions. 2006. Issue 8. Aug. P. 707—711.

8. Maltsev A., Pestretsov V., Maslennikov R., Khoryaev A. Triangular systolic array with reduced latency for QR-decomposition of complex matrices. Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium. 2006. P. 602—608.
9. Karcev M. A., Brik V. A. Vychislitel'nye sistemy i sinhronnaja arifmetika. M.: Radio i svjaz', 1981. 359 p.
10. Karcev M. A. Arifmetika cifrovych mashin. M.: Nauka, 1979. 575 p.
11. Ostanevich A. V. Jeksperimental'noe issledovanie podderzhki predmetnyh vychislenij v arhitekture s javno vyrashennym paralelilzom. Informacionnye tehnologii i vychislitel'nye sistemy. 1999. N. 1.
12. Kostin A. E., Shanguin V. F. Organizacija i obrabotka struktur dannyh v vychislitel'nyh sistemah. M.: Vysshaja shkola, 1987. 245 p.
13. Hegde G., Prasanna C., Raj P., Vaya P. R. Implementation of Systolic Array Architecture for Full Search Block Matching Algorithm on FPGA. European Journal of Scientific Research. 2009. Vol. 33, N. 4. P. 606—616.
14. Knjaz'kov V. S., Biktashev R. A. Parallel'nye vychislitel'nye sistemy s OKMD-arhitekuroj: uch. posobie. Penza: RIO Penzenskogo politehnicheskogo in-ta, 1991. 51 p.
15. Fedjunin R. N. Sposob realizacii apparatnogo sloja VLIW arhitekтуры na baze sistolicheskikh struktur. Izvestia Vysshih Uchebnyh Zavedenij Povolzhskij Region (Tehnicheskije Nauki), 2013. N. 2. P. 15—22.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ INFORMATION TECHNOLOGIES IN EDUCATION

УДК [378:004]:004.8

Л. Р. Черняховская, д-р техн. наук, проф., lrchern@yandex.ru,

И. Б. Герасимова, д-р техн. наук, доц., tarot_gera@mail.ru,

А. Р. Давлетбаева, salavatova.a@bk.ru

ФГБОУ ВПО "Уфимский государственный авиационный технический университет" (УГАТУ), Уфа

Разработка системы нечеткого логического вывода для анализа результативности процесса дистанционного обучения

На процесс дистанционного обучения оказывают влияние различные факторы, что обуславливает сложность анализа качества данного процесса и прогнозирования его результатов. Проведен системный анализ указанных факторов на основе триадного подхода, а также сделана попытка оценить это влияние с использованием системы нечеткого логического вывода.

Ключевые слова: дистанционное обучение, системный подход, триада, система нечеткого логического вывода

Введение

Применение дистанционных технологий в образовательной системе позволяет человеку не только эффективно использовать свое время для выполнения учебных и профессиональных задач, но и предоставляет возможности для свободного плодотворного сотрудничества в рамках создаваемой информационной среды, обусловленного потребностью личности в самореализации, в раскрытии своего профессионального потенциала и творческих способностей. Обучение с применением дистанционных образовательных технологий позволяет человеку более свободно выражать себя, преодолевая информационные, коммуникационные и временные барьеры.

На процесс дистанционного обучения (ДО) вследствие специфических особенностей его реализации, а также имеющихся в процессе ограничений оказывает влияние множество различных факторов. В данной работе рассмотрена системная модель факторов, влияющих на результативность дистанционного обучения, а также выполнена оценка этого влияния с помощью методики, основанной на применении системы нечеткого логического вывода.

Системная модель факторов, влияющих на результативность процесса дистанционного обучения

Для анализа факторов, оказывающих влияние на процесс дистанционного обучения, можно использовать системную модель, представляющую собой иерархически обусловленное соединение триад (рис. 1). Под триадой будем понимать соединение

и взаимодействие трех элементов (объектов, факторов), образующих единое целое, т. е. триада рассматривается как элементарная система [1].

Рассмотрим отношения, определенные на каждом уровне представленной системной модели факторов.

Уровень изучаемой дисциплины отражает выбранный студентом дистанционный курс.

Уровень факторов предполагает разделение оказывающих на процесс дистанционного обучения факторов на внешние и внутренние. К внешним факторам в отношении предлагаемого учебного курса можно отнести его длительность, новизну, актуальность, интенсивность, объем изучаемой информации, а также состояние окружающего пространства (цвет, форма и звук). К внутренним факторам относятся потребности и ожидания обу-



Рис. 1. Системная модель факторов, влияющих на результативность ДО

чающегося, его личностные и психологические особенности, эмоциональное состояние, личный и профессиональный опыт, уровень имеющихся знаний, умений и навыков (ЗУН) и др. [2].

На *уровне участников* результативность процесса дистанционного обучения предлагается рассматривать с точки зрения основных действующих лиц данного процесса. Одним из главных факторов, оказывающих положительное влияние на результативность студента в дистанционном обучении, является мотивация, так как только при высоком уровне мотивации можно говорить о готовности студента к обучению. Однако этот фактор не является определяющим, так как студенты могут записаться на большое число курсов, но не приступить к их изучению, например, вследствие недостатка времени. Поэтому наличие времени и умение его эффективно распределять, а также такие качества личности как самостоятельность, ответственность, усидчивость, целеустремленность и дисциплинированность можно отнести к факторам, оказывающим положительное влияние на процесс дистанционного обучения. Препятствиями в успешном освоении дистанционного учебного курса для студента могут стать: отсутствие доступа к сети Интернет, неуверенность в себе, изолированность, страхи, скука, отсутствие живого общения и др.

Преподаватель, или тьютор, является второй ключевой фигурой в системе дистанционного обучения, так как именно он организует весь процесс передачи информации и знаний в том объеме и ритме, который соответствует потребностям студента. Основная задача преподавателя заключается в правильном выборе ритма проведения занятий и оптимальной нагрузки, в регулярном проведении консультаций и процедур контроля получаемых знаний. Кроме этого, преподаватель при создании и использовании дистанционных курсов должен учитывать ряд требований к содержанию и изложению учебного материала, таких как логичность, последовательность, наглядность, простота, разнообразие видов предлагаемой деятельности и др.

Еще одним не менее важным участником процесса дистанционного обучения является координатор (специалист поддержки), задача которого состоит в отслеживании работы всей системы дистанционного обучения, а также в оказании необходимой поддержки при возникновении каких-либо трудностей. Эффективное взаимодействие всех трех элементов обеспечивает целостность всего процесса обучения.

Уровень организации мышления — это следующий уровень. В результате самостоятельного изучения первоисточников, учебного материала, выполнения заданий, регулярных обсуждений и участия в интернет-конференциях и вебинарах у студента формируются навыки анализа, синтеза, системного и творческого мышления, которые в дальнейшем активно используются в различных сферах его жизнедеятельности.

Уровень форм развития позволяет рассмотреть эффективность процесса дистанционного обучения с точки зрения его результатов, таких как формирование у студента новых знаний, умений и навыков, повышение его профессионализма и открытие новых перспектив для развития личности.

Представление влияющих на результативность дистанционного обучения факторов на основе триадного подхода позволяет привести их в единую сложную систему, раскрыть их взаимовлияние друг на друга как с позиции иерархии (уровней), так и с позиции множества элементов. Такое целостное системное виденье процесса обучения помогает как студенту, так и преподавателю обратить внимание на те факторы, которые не учитываются или остаются без внимания.

Представление системной модели факторов, влияющих на результативность процесса дистанционного обучения, в виде соединения триад дает возможность рассмотреть взаимовлияние факторов, находящихся на различных уровнях иерархии. Рассмотрим внутренние и внешние факторы, влияющие на результативность дистанционного обучения на уровнях изучаемой дисциплины и участников рассматриваемого процесса, на основе анализа триады "Студент—Учебный курс ДО—Преподаватель" (рис. 2).

Проблема оценки влияния факторов, обусловленных элементами выделенной триады, отличается достаточно высоким уровнем неопределенности, которая выражается, в том числе, в необходимости учета так называемого "человеческого фактора" — личностных и психологических особенностей участников процесса дистанционного обучения, а также характеристик предоставляемого в учебном курсе материала, которые обычно достаточно сложно измерить.

При описании и решении задач подобного типа зачастую используется подход на основе теории нечетких множеств и аппарата нечеткой логики, позволяющий описывать качественные, неточные понятия и наши знания об окружающем мире, а также оперировать этими знаниями в целях получения новой информации. Нечеткий вывод занимает центральное место в нечеткой логике и системах нечеткого управления и представляет собой процедуру или алгоритм получения заключений на основе нечетких условий или предпосылок с использованием понятий нечеткой логики. Системы нечеткого вывода предназначены для реализации процесса нечеткого вывода и служат концептуальным базисом всей современной нечеткой логики [3].

Для решения задачи оценки результативности дистанционного обучения в контексте триады "Студент—Учебный курс ДО—Преподаватель" предлагается использовать систему нечеткого логического вывода (СНВ), основанную на алгоритме нечеткого вывода Такаги—Сугено—Канга (*Takagi—Sugeno—Kang*). База знаний в модели типа Сугено является

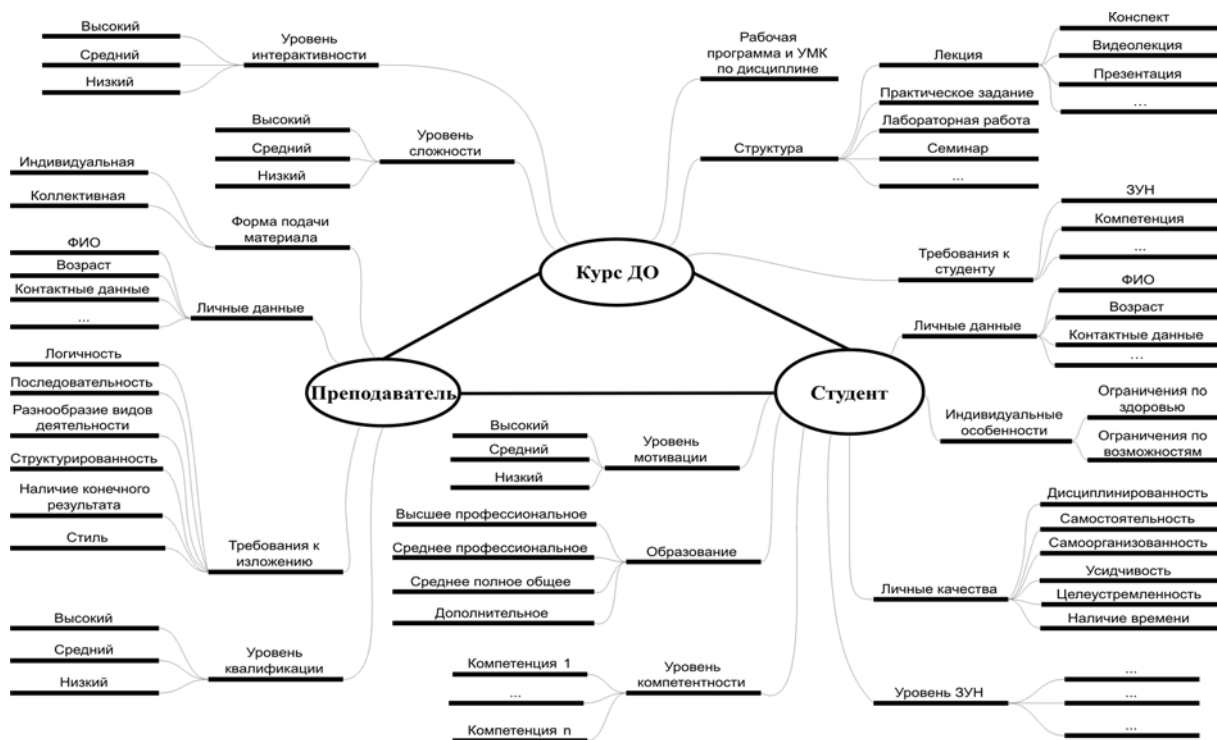


Рис. 2. Ментальная карта взаимодействия участников процесса дистанционного обучения

гибридной — ее правила содержат посылки в виде нечетких множеств и заключения в виде четкой линейной функции. К разрабатываемой базе знаний предъявляются следующие требования: семантическая целостность, непротиворечивость, полнота и непрерывность знаний, а также корректность взаимодействия пользователя с интеллектуальной системой. Таким образом, в процессе разработки базы знаний важно минимизировать ошибки, связанные с представлением экспертных знаний, толкованием их смысла, формой представления, а также с разработкой машины логического вывода и организацией процесса взаимодействия пользователя с интеллектуальной системой [4].

На начальном шаге создания системы нечеткого вывода были проведены анализ и выборка факторов, оказывающих наибольшее влияние на результативность дистанционного обучения в рамках выделенной триады. Выбор наиболее влиятельных факторов был проведен на основе метода экспертных оценок, однако для данной цели может быть также использован подход, основанный на применении нечетких когнитивных карт [5]. Таким образом, СНВ оценки результативности курса дистанционного обучения можно рассматривать как двухуровневую иерархическую систему, нижний уровень которой представлен тремя подсистемами нечеткого вывода, решающими задачи оценки профессиональных качеств преподавателя, качества дистанционного учебного курса и личных качеств студента. Верхний уровень создаваемой СНВ непосредственно решает задачу оценки результативности дистанционного обучения для студента на основе вы-

ходных данных подсистем нижнего уровня. Модель СНВ оценки результативности курса дистанционного обучения для студента приведена на рис. 3.

Рассмотрим входные переменные нижнего уровня создаваемой СНВ более подробно.

Для подсистемы нечеткого логического вывода, решающей задачу оценки профессиональных качеств преподавателя, входными параметрами являются следующие.

- **Квалификация.** Эта переменная позволяет учесть квалификацию преподавателя: ученую степень, звание, преподаваемые дисциплины, пройденные программы повышения квалификации и др. Значения для данной входной переменной будем определять экспертным методом на основе анализа предоставленной о преподавателе информации.
- **Научно-педагогический стаж.** Данная переменная позволяет учесть суммарную продолжительность трудовой деятельности в образовательных



Рис. 3. Модель СНВ оценки результативности курса дистанционного обучения для студента

учреждениях, связанную с учебным процессом, область научных исследований, участие в грантах и т. д. [6].

- **Количество публикаций.** Определяет количество трудов (статей, пособий, монографий и т. д.), опубликованных преподавателем за установленный период времени. Данный параметр отражает научную активность преподавателя, уровень его компетенций в профессиональной области.

Для подсистемы нечеткого логического вывода, решающей задачу оценки качества дистанционного курса, входные параметры такие.

- **Степень соответствия требованиям РП и УМК.** Параметр позволяет оценить соответствие контента дистанционного курса требованиям рабочей программы и УМК по изучаемой дисциплине. Значение данного параметра можно представить как степень присутствия в созданном дистанционном курсе требуемых материалов и инструментов обучения.
- **Степень соответствия требованиям к изложению материала.** К требованиям, предъявляемым к содержанию учебного материала курса, можно отнести структурированность, логичность, последовательность, простоту, разнообразие видов предлагаемой деятельности и т. д. Значения для данной входной переменной будем определять экспертным методом на основе анализа содержания курса.
- **Уровень интерактивности.** Данный параметр характеризует степень применения в дистанционном курсе интерактивных технологий, позволяющих студентам в активном режиме взаимодействовать с обучающей информационно-коммуникационной системой.

Значения для данных входных переменных подсистемы нечеткого логического вывода, решающей задачу оценки личных качеств студента — участника курса дистанционного обучения, предлагается получать на основе соответствующих процедур тестирования и самообследования, проводимых перед началом курса.

- **Уровень мотивации.** Данный параметр позволяет учесть уровень мотивации студента, который в значительной степени определяет эмоционально-оценочные реакции и формирует predisposition к тем или иным действиям и поступкам.
- **Уровень имеющихся ЗУН.** Параметр позволяет оценить степень наличия у студента знаний, умений и навыков (ЗУН), имеющих значение для освоения предлагаемого дистанционного курса.
- **Оценка личных качеств.** Данный параметр предполагает учет таких личностных и психологических характеристик студента, как наличие свободного времени, самостоятельность, целеустремленность, дисциплинированность, ответственность и др.

Для всех термов входных переменных определены симметричные функции принадлежности, фор-

мируемые с использованием гауссовского распределения.

Следующим этапом являлась разработка базы правил для СНВ в форме условных высказываний типа "ЕСЛИ—ТО", представляющих нечеткую импликацию:

$$\begin{aligned} \text{ЕСЛИ } x_{i1} \text{ есть } A_{i1} \text{ И } x_{i2} \text{ есть } A_{i2} \text{ И } x_{in} \text{ есть } A_{in} \\ \text{ТО } y_i \text{ есть } B_i. \end{aligned} \quad (1)$$

Здесь A_{ij} , B_i — лингвистические значения, идентифицированные нечетким способом через соответствующие функции принадлежности $\mu A(x_i)$ и $\mu B(y)$ для переменных x и y . Для всех термов переменных, входящих в правила типа (1), определены гауссовы функции принадлежности:

$$A_{ij}(x) = \exp\left[-\frac{1}{2}\left(\frac{x - a_{ij}}{\beta_{ij}}\right)^2\right], \quad B_i(y) = \exp\left[-\frac{1}{2}\left(\frac{y - a_i}{\beta_i}\right)^2\right].$$

Примеры правил нечеткого вывода созданной СНВ могут быть представлены в следующем виде:

- **ЕСЛИ** квалификация *средняя* **И** педагогический стаж *средний* **И** количество публикаций *высокое* **ТО** профессиональные качества преподавателя *средние*;
- **ЕСЛИ** степень соответствия требованиям РП и УМК *высокая* **И** уровень интерактивности *низкий* **И** степень соответствия требованиям к изложению материала *средняя* **ТО** качество курса *ДО среднее*;
- **ЕСЛИ** мотивация *высокая* **И** уровень имеющихся ЗУН *низкий* **И** оценка личных качеств студента *высокая* **ТО** личные качества студента *высокие*;
- **ЕСЛИ** профессиональные качества преподавателя *средние* **И** качество курса *ДО среднее* **И** личные качества студента *высокие* **ТО** результативность курса *ДО* для студента *средняя*.

На основе экспертных знаний было сформулировано более 200 правил. На рис. 4 (см. вторую сторону обложки) приведена визуализация поверхности системы нечеткого логического вывода, отражающая зависимость выходной переменной верхнего уровня *Результативность курса ДО* от входных переменных *Профессиональные качества преподавателя* и *Личные качества студента*.

Проанализировав полученную поверхность СНВ, можно сделать вывод о том, что она отвечает экспертным представлениям в рассматриваемой предметной области. Так, например, можно увидеть, что чем выше профессиональные качества преподавателя и лучше личные качества студента, тем с большей уверенностью можно говорить о результативности предлагаемого курса дистанционного обучения. В то же время можно заметить, что личные качества студента оказывают на результативность дистанционного обучения более сильное влияние.

Дальнейшие исследования будут связаны с разработкой на основе созданной СНВ соответствующей нейронечеткой сети, которая позволит учиты-

вать существующую информацию о процессе дистанционного обучения и его результатах для обучения сети. Адаптивная система нейронечеткого вывода позволит определять параметры функций принадлежности, которые лучше всего соответствуют созданной СНВ, а также формировать правила нечеткой продукции путем извлечения скрытых закономерностей из данных обучающей выборки.

Заключение

В результате проведенного исследования были рассмотрены и проанализированы факторы, оказывающие влияние на процесс дистанционного обучения. Была предложена системная модель представления факторов, влияющих на результативность процесса дистанционного обучения. Данная модель позволяет привести все факторы в единую сложную систему и раскрыть их влияние друг на друга. Кроме этого, в статье была предложена методика оценки внутренних и внешних факторов, влияющих на результативность дистанционного обучения на уровнях изучаемой дисциплины и участников рассмат-

риваемого процесса (в рамках анализа триады "Студент—Учебный курс ДО—Преподаватель"), на основе теории нечетких множеств с помощью построения системы нечеткого логического вывода.

Список литературы

1. **Гузайров М. Б., Ильясов Б. Г., Герасимова И. Б.** Системный подход к анализу сложных систем и процессов на основе триад // Проблемы управления. 2007. № 5. С. 32—38.
2. **Гузайров М. Б., Черняховская Л. Р., Герасимова И. Б.** и др. Анализ и управление качеством образовательных процессов на основе знаний // Труды IX Международной конференции "Проблемы управления и моделирования в сложных системах". Самара, 2007. С. 321—327.
3. **Леоненков А. В.** Нечеткое моделирование в среде MATLAB и fuzzyTECH. СПб: БХВ-Петербург, 2003. 736 с.
4. **Черняховская Л. Р., Старцева Е. Б., Муксимов П. В.** и др. Поддержка принятия решений при стратегическом управлении предприятием на основе инженерии знаний / Под ред. Л. Р. Черняховской. Уфа: АН РБ, Гилем, 2010. 128 с.
5. **Черняховская Л. Р., Герасимова И. Б., Салаватова А. Р.** и др. Оценка влияния социально-психологических факторов на качество подготовки студента с применением нечетких когнитивных карт // Вестник УГАТУ. 2014. Т. 18.
6. **Письмо** Минобрнауки РФ № АФ-947/96 от 26.10.2004 "О размерах и условиях оплаты труда работников образовательных учреждений в 2005 году" // Вестник образования. 2005. № 2.

L. R. Chernyakhovskaya, Dr. Sci. Tech., Professor of Technical Cybernetics Department,
e-mail: lrchern@yandex.ru;

I. B. Gerasimova, Dr. Sci. Tech., Professor of Automated Control and Management Systems Department,
e-mail: tarot_gera@mail.ru;

A. R. Davletbaeva, Graduate Student, e-mail: salomeia07@mail.ru
Ufa State Aviation Technical University (USATU), Ufa

Development of Fuzzy Inference System for Distance Learning Process Analysis

Currently distance learning technology (d-learning or e-learning) is an effective way to support educational process which has many advantages. Distance learning is influenced by number of various factors, therefore it's quality analysis and outcome prediction opportunity tasks have fairly high level of complexity and uncertainty. In this article we propose to provide systems analysis of these factors on the basis of triadic approach and attempt to estimate their effect on distance education process efficiency using fuzzy inference system. Triadic approach allows to bring all factors into a unified complex system and reveal their influence on each other. Fuzzy inference system makes it possible to estimate impacting of internal and external factors affecting distance learning on the levels of discipline and education process participants.

Keywords: distance learning, e-learning, systems analysis, triadic approach, Fuzzy Inference System

References

1. **Guzairov M. B., Ilyasov B. G., Gerasimova I. B.** Sistemyi podkhod k analizu slozhnykh sistem i processov na osnove triad. *Problemy upravleniya*. 2007. N. 5. P. 32—38 (in Russian)
2. **Guzairov M. B., Chernyakhovskaya L. R., Gerasimova I. B., Nugaeva K. R.** *Trudy IX Mezhdunarodnoi konferentsii "Problemy upravleniya i modelirovaniya v slozhnykh sistemakh"*, Samara, 2007. P. 321—327 (in Russian)
3. **Leonenkov A. V.** *Nechetkoe modelirovanie v srede MATLAB i fuzzyTECH* (Fuzzy modeling in MATLAB and fuzzyTECH). Saint Petersburg: BKhV-Peterburg, 2003. 736 p. (in Russian)
4. **Chernyakhovskaya L. R., Startseva E. B., Muksimov P. V., Makarov P. V., Malahova A. I.** *Podderzhka prinyatiya reshenii pri*

strategicheskoy upravlenii predpriyatiem na osnove inzhenerii znaniy (Decision making support at the strategic management of the enterprise on the basis of knowledge engineering). Ufa: AN RB, Gilem, 2010. 128 p. (in Russian)

5. **Chernyakhovskaya L. R., Gerasimova I. B., Salavatova A. R., Mukhamedrakhimova L. N.** *Vestnik UGATU*, 2014. Vol. 18, N. 4(65), P. 134—141. (in Russian)

6. **O razmerakh i usloviyakh oplaty truda rabotnikov obrazovatel'nykh uchrezhdenii v 2005 godu.** *Pis'mo Minobrnauki RF* от 26.10.2004 N AF-947/96 (About the size and terms of educational institutions employees remuneration in 2005. Letter of the Russian Federation Ministry of Education from 26.10.2004. N AF-947/96). Moscow, *Vestnik obrazovaniya*. 2005. N. 2 (in Russian)

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В БИМЕДИЦИНСКИХ СИСТЕМАХ INFORMATION TECHNOLOGIES IN BIOMEDICAL SYSTEMS

УДК 57.087:57.023

М. С. Захаров, e-mail: sergei101@mail.ru
Начальник управления ЗАО "Сбербанк—Технологии", г. Москва

Динамика во времени вегетативных показателей на основе компьютерного спектрально-временного анализа ритмограмм сердца

Представлена методология компьютерного спектрально-временного анализа кардиоинтервалов, позволяющая следить за состоянием вегетативной нервной системы (ВНС) человека во времени. Основу предложенной методологии составляет спектральный анализ Фурье "оконного типа". Использование преобразований Фурье данного типа с уменьшенным числом отсчетов дает возможность исследовать нестационарные спектральные свойства ритмограмм сердца. Изучены мощностные спектральные характеристики, относящиеся к различным отделам ВНС. По площадям спектральных кривых, относящихся к выделенным выше областям спектра, можно судить о балансе механизмов расхода и восстановления энергии в организме и адаптационных процессах во времени.

Ключевые слова: фотоплетизмограмма, кардиоинтервалы, ритмограмма, преобразования Фурье "оконного типа"

Введение

Как известно, динамика временных рядов кардиоинтервалов (КИ) составляет основу метода вариабельности сердечного ритма (ВСР), получившего широкое распространение в донозологической диагностике при исследованиях состояний организма на грани нормы и патологии [1–6]. Данный метод начал развиваться еще в 60-е годы в СССР и до сегодняшнего дня активно используется при подготовке космонавтов. Анализ ВСР по сути представляет собой статистический анализ и позволяет проводить оценку регуляторных систем организма, в частности, функционального состояния различных отделов вегетативной нервной системы (ВНС).

С временными свойствами кардиоинтервалов тесно связаны спектральные особенности. Традиционный спектральный анализ Фурье представляет собой инструмент исследования так называемых стационарных случайных процессов. Данный подход означает, что рассматриваемые процессы можно приближенно считать эргодическими. На самом деле допустимость такого предположения довольно трудно обосновать, поскольку живой организм постоянно испытывает процессы адаптации к изменяющимся условиям внешней среды. Действительно, естественным проявлением жизни являются не квазистационарные, а переходные процессы. По-видимому, методологически неверно рассматривать ВСР вне времени. Поскольку сердечный ритм реализуется во времени, то и основные показатели ВНС также необходимо анализировать динамически.

В настоящей работе предложена методология исследования динамических спектрально-временных свойств кардиоинтервалов, позволяющая исследовать вегетативные функции нервной системы и процесс адаптации организма во времени. Показана существенная нестационарность, возникающая в системах управления сердечным ритмом. Исследованы мощностные характеристики процессов, позволяющие установить временную динамику различных отделов ВНС и ответить на вопрос, как происходит процесс адаптации к изменяющимся условиям внешней среды.

Особенности экспериментального снятия ритмограмм методом фотоплетизмографии

Фотоплетизмограммы (ФПГ) в данной работе регистрировались с помощью аппаратно-программного комплекса (АПК) "Биомышь", разработанного компанией "Нейролаб" (Москва) и предназначенного для проведения экспресс-оценки функционального состояния человека. Основу АПК составлял фотоэлектронный датчик, представляющий собой оптоэлектронную пару (ИК светодиод и фотоприемник) (рис. 1).

Сигнал с датчика подавался на селективный усилитель и оцифровывался с частотой дискретизации 1 кГц. Программное обеспечение позволяло проводить не только математическую обработку по методике ВСР, но также анализировать временную форму сигнала ФПГ.

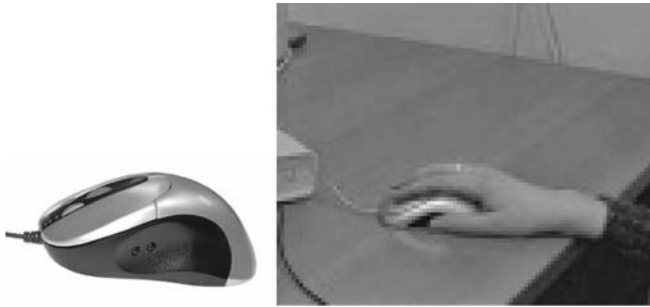


Рис. 1. АПК "Биомышь индивидуальная" (ЗАО Нейролаб) [7]

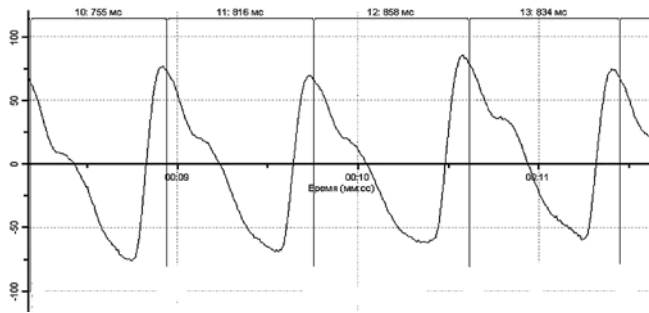


Рис. 2. Временной отрезок сигнала ФПГ, снятый с АПК "Биомышь"

Длительности кардиоинтервалов (мс), соответствующие 100 отсчетам

Номера отсчетов	1	2	3	4	5	6	7	8	9	10
0	828	846	893	875	831	837	830	795	747	755
10	816	858	834	832	863	852	815	810	831	796
20	675	833	858	859	815	812	848	816	762	802
30	856	854	842	831	844	845	811	776	802	782
40	777	791	827	854	844	847	880	884	819	782
50	797	828	809	791	794	765	696	704	726	721
60	749	804	805	784	804	860	852	808	793	834
70	818	819	817	862	854	828	819	827	812	801
80	795	814	837	793	807	820	850	813	822	828
90	821	801	817	821	846	821	837	828	833	809

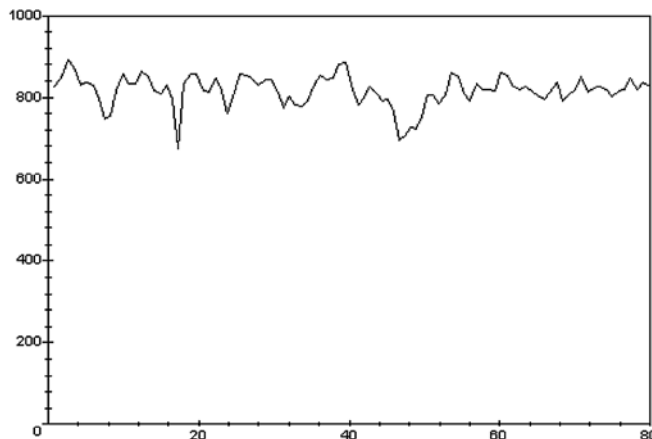


Рис. 3. Ритмограмма (зависимость длительностей кардиоинтервалов в мс, как функция времени, с), записанная по 100 сердечным ритмам. Здесь среднее значение длительности кардиоинтервалов равно $\langle \Delta t_k \rangle = 815,7$ мс (пульс ~ 74 уд/мин)

Временная форма сигнала ФПГ, относящаяся к здоровому пациенту и снятая с датчика АПК "Биомышь", показана на рис. 2.

Длительности временного ряда кардиоинтервалов выделялись из фотоплетизмограммы программным образом по максимумам отсчетов ФПГ и заносились в таблицу. При этом приборная погрешность определения длительности кардиоинтервала составляла ~ 1 мс.

Набор длительностей соответствующих кардиоинтервалов позволяет получить временное их представление или зависимость от текущего времени и построить ритмограмму. Пример таким образом полученной ритмограммы здорового пациента показан на рис. 3, на котором видно, как проявляется синусный характер сердечного ритма.

Компьютерный спектрально-временной анализ ритмограмм на основе преобразования Фурье "оконного" типа

В дальнейшем необходимо отметить, что все вычисления выполнены в среде "Maple 5v4".

Для удобства сравнения различных ритмограмм во времени необходимо их привести к эквидистантному временному интервалу. Это можно сделать путем сплайновой кубической интерполяции.

Исходный Фурье спектр будем вычислять по формуле

$$F(\nu) = \sum_{k=1}^N \exp(2\pi i \nu t_k) (\Delta t_k - \langle \Delta t_k \rangle) \Delta t_k; \quad (1)$$

$$P(\nu) = \frac{|F(\nu)|^2}{2T}, \quad (2)$$

где $P(\nu)$ — спектральная плотность мощности; T — временной интервал (с), соответствующий 100 сердечным циклам ($N = 100$); Δt_k — длительности кардиоинтервалов (мс), t_k — текущее время (с):

$$T = \sum_{k=1}^N \Delta t_k; \quad (3)$$

$$t_k = \sum_{l=1}^k \Delta t_l. \quad (4)$$

Для спектра Фурье, вычисленного по эквидистантным временным отсчетам, будем использовать следующее выражение [8]:

$$F(\nu) = \sum_{k=-N/2}^{N/2} \exp(2\pi i \nu t_{k+N/2}) (x_{k+N/2} - \langle x \rangle) \Delta t, \quad (5)$$

где x_k — временные отсчеты на интервалах интерполяции $t_k = k\Delta t = kT/N$ ($k = 0, 1, \dots, N$; $T = 80$ с; $N = 100$).

Для сравнения на рис. 4 представлены спектрограммы Фурье исходного массива кардиоинтервалов и массива, полученного интерполяцией на эквидистантных временных отрезках.

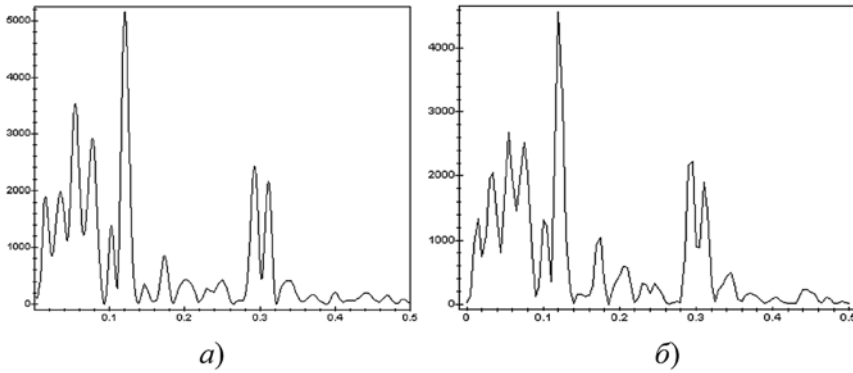


Рис. 4. Фурье-спектры ритмограмм исходной (а) и полученной сплайновой кубической интерполяцией (б). Спектральное разрешение $\sim 0,013$ Гц. По оси абсцисс — частота, Гц; по оси ординат — спектральная плотность мощности, $(мс)^2/Гц$

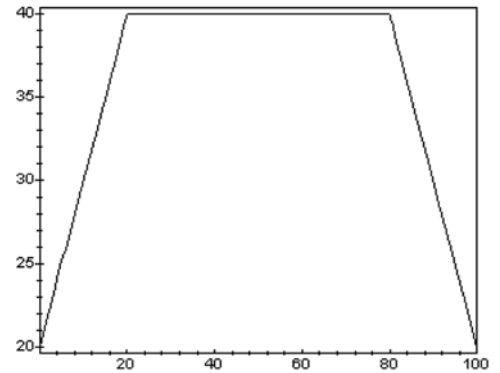


Рис. 5. Вид функции $W(n)$, описывающей зависимость числа временных точек суммирования в выражении (6) от времени ($N = 100$, $M = 40$)

Основная идея использования спектров "оконного" типа состоит в сознательном уменьшении числа временных точек N , необходимых для построения спектра, до величины $M < N$ [8]. При этом, естественно, уменьшается спектральное разрешение (в N/M число раз), но зато появляется возможность анализа спектрограмм во времени. Тогда выражение (5) видоизменится:

$$F(v, t_k) = \sum_{k=-M/2}^{M/2} \exp(2\pi i v t_{k+m}) \times (x_{k+m} - \langle x \rangle) \Delta t. \quad (6)$$

Пределы суммирования в выражении (6) справедливы для временных точек $M/2 < k < N - M/2$. Таким образом, возникает возможность временного анализа ритмограмм в интервале времен $(N - M)\Delta t$. Формально можно расширить временной интервал до полного (от 0 до $N\Delta t = T$) введением функции $W(n)$, описывающей зависимость числа временных точек от времени. Ее вид для $N = 100$ и $M = 40$ представлен на рис. 5.

Отметим, что в последнее время активно развивается направление, связанное с использованием вейвлет-преобразований. В качестве примера можно привести работу [9], в которой исследовались вейвлет-спектры кардиоинтервалов. Помимо неоднозначного выбора базиса спектры такого типа обладают отличительным свойством зависимости спектрального разрешения от частоты. Для наших целей более удобным оказалось использование "оконного" преобразования Фурье с независимым спектральным разрешением.

Пример спектров "оконного" типа, рассчитанных в первые пять секунд, показан на рис. 6.

Из рис. 6 следует, что динамика спектров проявляется даже на небольших временных интервалах.

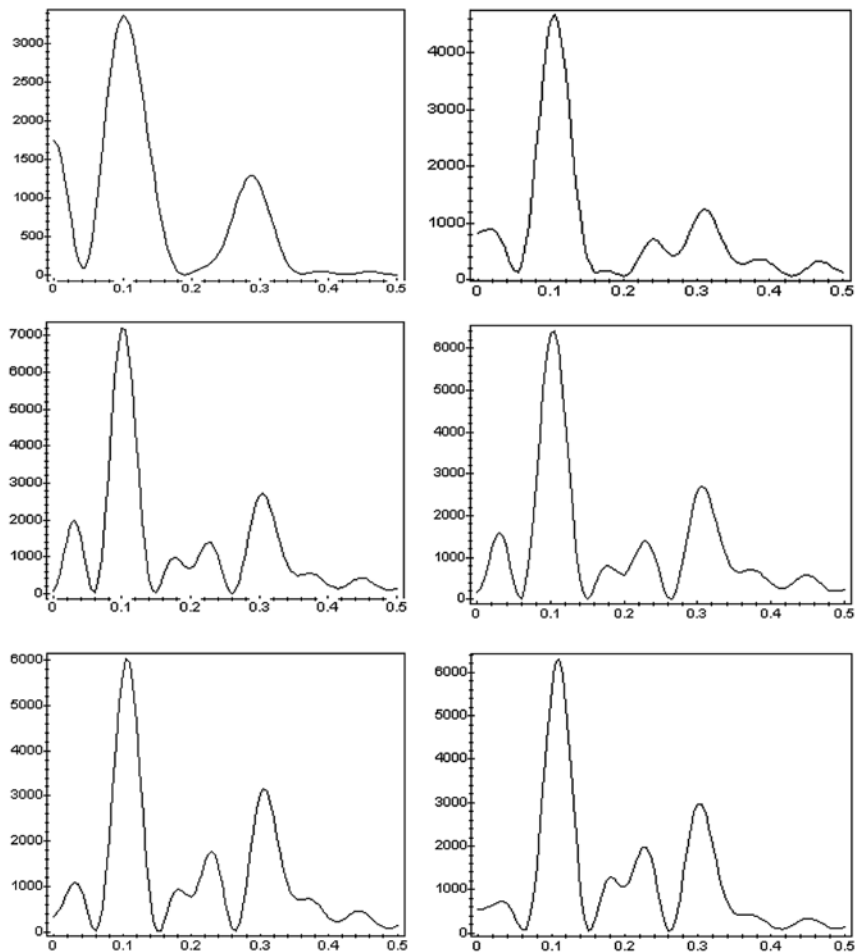


Рис. 6. Динамика спектра Фурье, рассчитанного по 40 из 100 временным отсчетам в течение первых пяти секунд (время увеличивается слева \rightarrow направо \rightarrow вниз). Спектральное разрешение ~ 0.06 Гц. По оси абсцисс — частота, Гц; по оси ординат — спектральная плотность мощности, $(мс)^2/Гц$

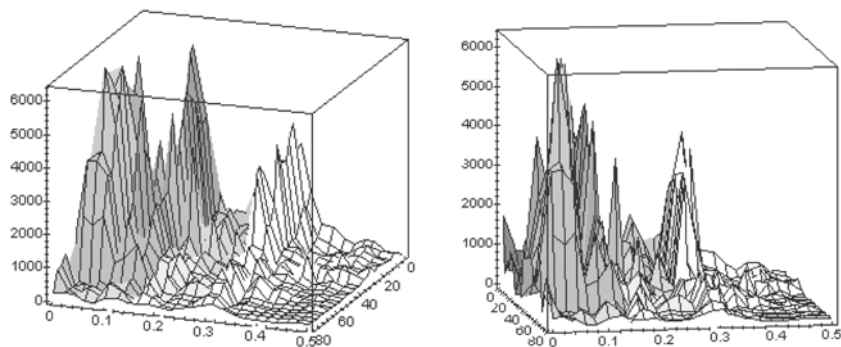


Рис. 7. Трехмерное представление спектров Фурье во времени в различных ракурсах. Ось 0–0,5 относится к частоте, Гц; 0–80 — ко времени, с

Здесь же приводится трехмерный вид спектров в различных ракурсах (рис. 7).

Исследование адапционных процессов организма методом спектрально-временного анализа ритмограмм сердца

Исследование процессов адаптации организма человека и животных является одним из фундаментальных направлений современной биологии и физиологии. Адаптационная деятельность обеспечивает не только выживание и эволюционное развитие, но и повседневное приспособление к изменениям окружающей среды. Г. Селье описывает фазовый характер адапционных реакций и обосновывает ведущую роль истощения регуляторных систем в развитии большинства патологических состояний и заболеваний. При этом система кровообращения может рассматриваться как чувствительный индикатор адапционных реакций организма в целом.

Определяющую роль в процессах адаптации и поддержании постоянства внутренней среды или установлении гомеостаза при различных воздействиях на организм играет вегетативная нервная система и взаимодействие ее отдельных систем. Как известно, симпатический отдел ВНС отвечает за мобилизацию внутренних ресурсов организма, за повышение энергообразования. Парасимпатический отдел ВНС отвечает за расслабление, отдых, сохранение и накопление жизненной энергии.

Применение спектрального анализа позволяет количественно оценить различные частотные области ритма сердца и наглядно представить соотношения разных составляющих, отражающих активность определенных отделов регуляторного механизма ВНС. Анализ спектральной плотности мощности колебаний дает информацию о распределении мощности в зависимости от частоты.

При спектральном анализе большое значение имеет объем анализируемой информации. При коротких записях (5 мин и менее) выделяют три главных спектральных области [4, 5]:

HF (High Frequency) — высокие частоты. Их частотный интервал составляет 0,15...0,4 Гц, а длительность — 2,5...6,6 с;

LF (Low Frequency) — низкие частоты. Частоты данной области относятся к интервалу 0,04...0,15 Гц, а длительности — 6,6...25,0 с;

VLF (Very Low Frequency) — очень низкие частоты или медленные волны. Их частоты — 0,015...0,04 Гц, а длительности — 25...66 с;

TF (Total Power TF = HF + LF + VLF) — полная, или суммарная, мощность во всех спектральных диапазонах.

До сих пор еще идут споры по поводу определения биологического значения и точных границ этих диапазонов, но в большинстве случаев исследователи сходятся на следующем понимании.

HF-диапазон отражает процессы парасимпатической активности, LF-диапазон связан с симпатической активностью, VLF-диапазон отражает гуморально-метаболические влияния. Гуморальная регуляция — один из эволюционно ранних механизмов регуляции процессов жизнедеятельности в организме, осуществляемый через жидкие среды организма (кровь, лимфу, тканевую жидкость) с помощью гормонов, выделяемых клетками, органами, тканями. У высокоразвитых животных и человека гуморальная регуляция подчинена нервной регуляции и составляет совместно с ней единую систему нейрогуморальной регуляции.

По площадям спектральных кривых, относящихся к выделенным выше областям спектра, судят о балансе механизмов расхода и восстановления энергии в организме. При выполнении неравенства $HF > LF$ происходит преобладание процесса восстановления энергетического потенциала. Неравенство $LF > HF$ означает преобладание процесса истощения энергетического потенциала. Нормой считается соотношение LF/HF (индекс вагосимпатического взаимодействия) $\sim 1,5...2$. Если данный коэффициент превышает 2, то это свидетельствует о том, что человек сильно мобилизован. Когда мобилизация происходит в момент выполнения какой-либо работы (нагрузки), то это нормально, а если человек все время находится во "взвинченном" состоянии, то ничего хорошего в этом нет. Другая крайность — сильное расслабление, когда соотношение $LF/HF < 1$. Это совершенно естественно для человека, который устал, и его организм переключился в режим отдыха.

Мощности выделенных областей спектра в абсолютных единицах ($мс^2$) и относительных HF/TF , LF/TF , VLF/TF (где $TF = HF + LF + VLF$) как функции времени показаны на рис. 8. Следует отметить возникающие во времени осцилляции, носящие не совсем регулярный характер и возникающие, по-видимому, из-за дискретного характера изменений в спектрах. Кроме того, из рис. 8, б видно, как относительные симпатическая и парасимпатическая регуляторные системы "работают" со временем в противофазе.

На рис. 8 можно проследить, как происходит адаптация сердечно-сосудистой системы организма в процессе снятия ФПГ. На участке 0...30 с парасимпатическая и симпатическая активности регуляторного механизма сбалансированы и находятся в относительном равновесии. Далее на участке 30...65 с возрастает роль гуморальных процессов и процессов активации симпатического звена регуляции. И лишь в промежутке времени 65...80 с наступает релаксация организма.

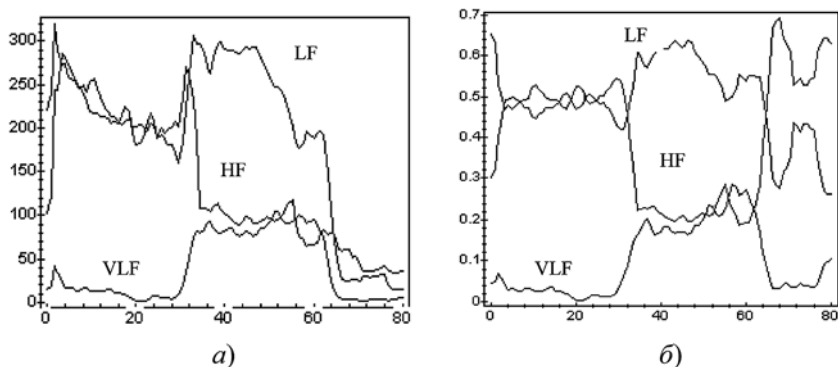


Рис. 8. Спектральные мощности как функции времени, относящиеся к областям HF, LF, VLF в абсолютных (мВт^2) — (а) и относительных единицах (в отношении к TF) — (б)

Централизация управления ритмом характеризуется индексом централизации IC, который вычисляется по формуле $IC = (HF + LF)/VLF$. Данный индекс возрастает при усилении автономных влияний. Суть его сводится к тому, что он отвечает на вопрос: справляется ли вегетативная нервная система (LF, HF) с функцией управления ритмом сердца, либо ей на помощь приходит сосудодвигательный (вазомоторный) центр и гуморальные механизмы регуляции (VLF). Неравенство $IC > 1$ означает, что процесс регуляции физиологических функций характеризуется преобладанием автономных влияний в управлении, отражая оптимальное функционирование системы. Если $IC < 1$, то процесс регуляции физиологических функций характеризуется преобладанием центральных влияний в управлении, отражая напряжение функционирования системы. График IC как функции времени показан на рис. 9. Здесь же приводится график вагосимпатического взаимодействия [5].

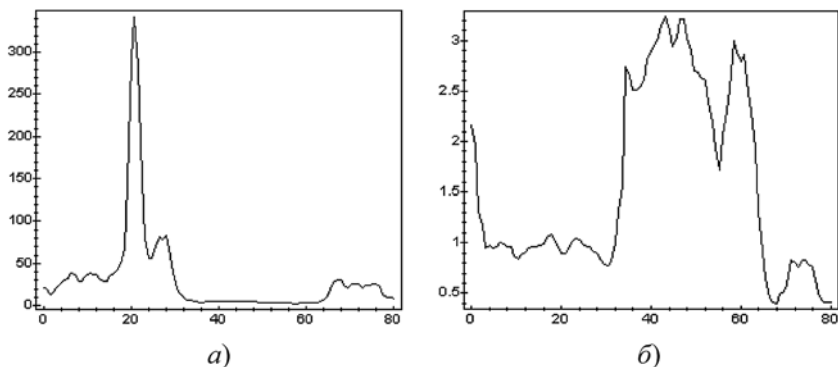


Рис. 9. Индексы централизации (а) и вагосимпатического взаимодействия (б) как функции времени, с

Представляет интерес также анализ максимального значения гармоники в той или иной области спектра [10]. На рис. 10 показаны зависимости во времени максимальной гармоники в областях HF (нижняя кривая) и LF (верхняя кривая) (а), а также в областях HF и VLF (б).

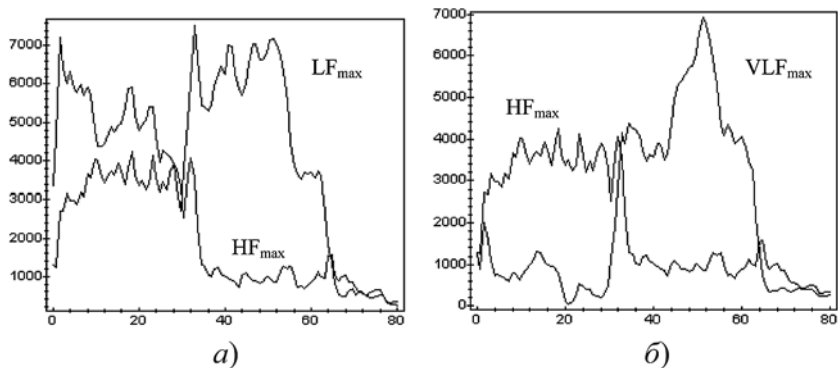


Рис. 10. Значения максимумов частот в различных областях (HF, LF) — (а) и (HF, VLF) — (б) как функции времени, с

Из графиков, изображенных на рис. 10, а, следует, что парасимпатическая система регуляции ритма сердца незначительно преобладает над симпатической на отрезке времени 0...30 с, но имеются и моменты времени, когда, наоборот, симпатическая преобладает над парасимпатической (моменты времени ~ от 30 до 65 с). Точно также из рис. 10, б следует, что парасимпатическая система главенствует во времени над системой VLF на отрезке 0...30 с, но также имеются моменты времени (область от 30 до 65 с) с обратным влиянием. Особенно это становится заметным тогда, когда максимумы различных спектральных областей приводятся на одном графике (рис. 11).

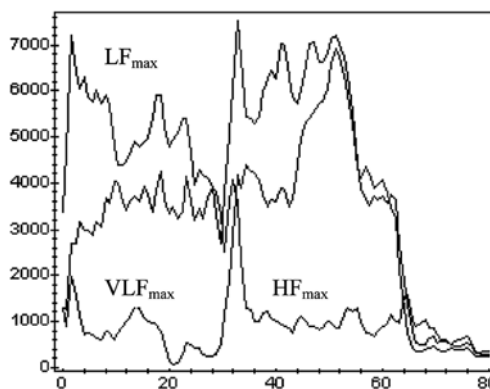


Рис. 11. Значения максимумов гармоник, относящихся к областям спектра HF, LF и VLF как функции времени, с

Результаты приведенного динамического спектрального анализа ритмограмм во времени свидетельствуют о существенной нестационарности механизмов регуляторного влияния на сердечный ритм. Дискретные изменения, возникающие в спектрах во времени, приводят к возникновению своеобразных осцилляций, носящих не вполне регулярный характер. Аналогичным образом проявляются зависимости во времени мощностных характеристик, относящихся к различным отделам нервногуморальной регуляции. Представленная методология позволяет проследить за тем, как происходит процесс адаптации организма во времени. В дальнейшем представляет интерес исследование динамики спектров и нестационарных процессов регуляции нервной системы на более протяженных временных отрезках, в частности, стандартном пятиминутном, а также при холтеровском мониторинге. Рассмотренный спектрально-временной анализ ритмограмм расширяет возможности использования метода ВСП в клинической практике и в донозологической диагностике.

Автор выражает признательность проф. Шилкову В. В. за полезные обсуждения полученных результатов.

1. **Баевский Р. М.** Прогнозирование состояний на грани нормы и патологии. М.: Медицина, 1979. 205 с.
2. **Баевский Р. М., Берсенева А. П.** Оценка адаптационных возможностей организма и риск развития заболеваний. М.: Медицина, 1997. 265 с.
3. **Баевский Р. М., Берсенева А. П.** Введение в донозологическую диагностику. М.: Слово, 2008. 220 с.
4. **Вариабельность** сердечного ритма. Стандарты измерения, физиологической интерпретации и клинического использования (Рекомендации Рабочей группы Европейского кардиологического общества и Северо-Американского общества стимуляции и электрофизиологии) // Вестник аритмологии. 1999. № 11. С. 52—77.
5. **Баевский Р. М., Иванов Г. Г., Чирейкин Л. В.** и др. Анализ variability сердечного ритма при использовании различных электрокардиографических систем. (Методические рекомендации) // Вестник аритмологии. 2001. № 24. С. 65—87.
6. **Михайлов В. М.** Variability ритма сердца: опыт практического применения метода. Изд. 2-е. Иваново: Иван. гос. мед. академия, 2002. 290 с.
7. **ЗАО "Нейролаб"**. Биомышь визуальная. URL: www.neurolab.ru/downloads/personal/biomouse_personal_booklet.pdf
8. **Бережной В. Н., Захаров М. С., Захаров С. М.** Динамика во времени спектральных особенностей сигналов пульсовой волны, полученных методом фотоплетизмографии // Биомедицинская радиоэлектроника. 2013. № 6. С. 36—43.
9. **Захаров С. М.** Вейвлет-анализ кардиоинтервалов // Биомедицинская радиоэлектроника. 2012. № 11. С. 43—48.
10. **Захаров М. С.** Временной спектральный анализ кардиоинтервалов // Сб. трудов XIV Международной заочной научно-практической конференции "Научная дискуссия: инновации в современном мире" (2 июля 2013 г.). М.: Междунар. центр науки и образования, 2013. С. 6—11.

M. S. Zakharov, e-mail: sergei101@mail.ru

Head of Department of CJSC "Sberbank—Technology", Moscow

Dynamics in Time Vegetative Indices Based on Computer Spectral Analysis of Heart Rhythmograms

The methodology of computing spectral temporal analysis of cardio intervals, which allows to monitor the state of the autonomic nervous system (ANS) a man in time is presented. The basis of the proposed methodology is the Fourier spectral analysis "window-type". Using Fourier transforms of this type with a reduced number of samples makes it possible to investigate the spectral properties of non-stationary cardio intervals. Rhythmogram heart rate was recorded by photoplethysmography at one hundred counts with using an optoelectronic sensor built in the computer mouse. For ease of comparison of various rhythmograms in time they all led to equidistant time interval. This procedure was carried out using a spline cubic interpolation. By singling out certain frequency ranges (HF — high frequency, LF — low frequency and VLF — very low frequency) the dynamics of the ANS of a healthy patient in time is studied. Power spectra characteristics relating to the various departments VNS is investigated. By area of spectral curves pertaining to selected regions of above spectrum, one can judge the balance of flow mechanisms and restore the body's energy and adaptation processes in time.

Keywords: photoplethysmogram, cardio interval, rhythmogram, "windowed" Fourier transforms

References

1. **Baevskiy R. M.** *Prognozirovaniye sostoyaniy na grani normy i patologii*. M.: Meditsina, 2009. 205 p.
2. **Baevskiy R. M., Berseneva A. P.** *Otsenka adaptatsionnykh vozmozhnostey organizma i risk razvitiya zabolevaniy*. M.: Meditsina, 1997. 265 p.
3. **Baevskiy R. M., Berseneva A. P.** *Vvedeniye v donozologicheskuyu diagnostiku*. M.: Slovo, 2008. 220 p.
4. **Variabel'nost'** serdechnogo ritma. Standarty lzmereniya, fiziologicheskoy interpretatsii i klinicheskogo ispol'zovaniya (Rekomendatsii Rabochey gruppy Evropeyskogo kardiologicheskogo obshchestva i Severo-Amerikanskogo obshchestva stimulyatsii i elektroфизиологии). *Vestnik Aritmologii*. 1999. N. 11. P. 52—77.
5. **Baevskiy R. M., Ivanov G. G., Chireykin L. V.** i dr. Analiz variabel'nosti serdechnogo ritma pri ispol'zovanii razlichnykh elektrokardiograficheskikh sistem. (Metodicheskie rekomendatsii). *Vestnik Aritmologii*. 2001. N. 24. P. 65—87.
6. **Mikhaylov V. M.** Variabel'nost' ritma serdtsa: opyt prakticheskogo primeneniya metoda. Izd. 2-e. Ivanovo: Ivan. Gos. med. akademiya, 2002. 290 p.
7. **ЗАО Neyrolab**. "Biomyshe' individual'naya". URL: www.neurolab.ru/downloads/personal/biomouse_personal_booklet.pdf
8. **Berezhnoy V. N., Zakharov M. S., Zakharov S. M.** Dinamika vo vremeni spektral'nykh osobennostey signalov pul'sovoy volny, poluchennykh metodom fotopletizmografii. *Biomeditsinskaya radioelektronika*. 2013. N. 6. P. 36—43.
9. **Zakharov S. M.** Veyvlet-analiz kardiointervalov. *Biomeditsinskaya radioelektronika*. 2012. N. 11. P. 43—48.
10. **Zakharov M. S.** Vremennoy spektral'nyy analiz kardiointervalov. Sb. trudov XIV Mezhdunarodnoy zaochnoy nauchno-prakticheskoy konferentsii "Nauchnaya diskussiya: innovatsii v sovremennom mire" (2 iyulya 2013 g.). M.: Mezhdunarodnyy tsentr nauki i obrazovaniya, 2013. P. 6—11.

Ю. А. Зак, д-р техн. наук, научный эксперт и консультант, Германия, Аахен,
e-mail: yuriy_zack@hotmail.com

Об одной задаче построения допустимых и оптимальных маршрутов доставки грузов

Рассмотрена важная для практических приложений задача построения оптимальных маршрутов движения грузовых транспортных средств (ТС) в условиях ограничений на сроки обслуживания пунктов и на грузоподъемность ТС. В качестве критериев оптимальности рассматриваются суммарные затраты на выполнение графика доставки грузов.

На основе установленных свойств допустимых и оптимальных решений задачи сформулированы правила отсева недопустимых и неоптимальных маршрутов и разработаны алгоритмы решения задачи модифицированными методами динамического программирования и ветвей и границ, которые проиллюстрированы числовым примером. Приведены результаты вычислительных экспериментов.

Ключевые слова: грузовые транспортные средства, оптимальные маршруты движения, ограничения на сроки обслуживания, грузоподъемность и объемы грузов, последовательные алгоритмы оптимизации, модифицированный метод ветвей и границ

Введение

Рассматриваются постановки, математические свойства и алгоритмы решения задач построения оптимальных маршрутов движения транспортных средств (ТС), обеспечивающих с наименьшими затратами времени и средств доставку грузов в установленных объемах и в определенные договорными обязательствами сроки. В литературе (см., например, [1–10]) задачи доставки однородного груза от некоторого множества производителей потребителям с использованием транспортного средства ограниченной вместимости, формулировались в виде математических моделей целочисленного линейного и булевого программирования, определения потоков в сетях, а также выбора кратчайшего пути на графе. Суммарные затраты на транспортировку груза между пунктами зависят от времени ТС в пути, объема транспортируемого груза и времени обслуживания пунктов. Сформулировано множество подобных задач, в постановке которых отражаются различные ограничения, диктуемые практикой. Классификация таких задач приведена, например, на сайте [5] и в работах [3, 4, 7]. Были рассмотрены алгоритмы решения сформулированных задач методами математического, динамического программирования или методом ветвей и границ [5, 6], которые не учитывали специфику конкретных математических моделей, либо эвристическими методами [7], позволяющими получить лишь приближенное решение сформулированной проблемы.

В отличие от известных публикаций автор рассматривает постановки задачи доставки однородных грузов, в которых предусмотрены ограничения на сроки доставки и загрузки груза, а также на ограниченную грузоподъемность ТС. Существенное отличие от других публикаций (см., например [5, 6]) заключается в наличии ограничений на допустимые времена обслуживания пунктов и в том, что при построении маршрутов для задач данного класса, если не ограничивать объемы грузов, которые нужно доставить или забрать в отдельных пунктах, необходимо учитывать, что в некоторые пункты ТС может заходить более одного раза, и следовательно, оптимальное решение задачи не может быть построено в виде циклического маршрута и замкнутого гамильтонового цикла. В работе приведены постановки задач, важных для практических приложений, исследуются их математические особенности. На основе установленных свойств допустимых и оптимальных маршрутов предлагаются эффективные алгоритмы их решения последовательными алгоритмами оптимизации, учитывающие специфику каждой конкретной задачи. Основное отличие предложенных в работе алгоритмов от рассматриваемых в публикациях [4–10] заключается в применении их к решению оригинальной, достаточно важной и сложной прикладной задачи. Автору не известны постановки и методы решения задачи теории расписаний в условиях одновременно действующих ограничений и на грузоподъемность транспортных средств (ТС), и на допустимые времена обслуживания пунктов, а также при необходимости из не-

которых пунктов забирать, а в некоторые пункты доставлять груз. В отличие от известной задачи коммивояжера, в общем случае, не существует решения задачи в виде замкнутого гамильтонового маршрута, так как ТС придется заезжать в некоторые пункты более, чем по одному разу. Также в некоторых случаях необходимы простои ТС до нижней границы допустимого времени обслуживания пункта. Для каждой вершины дерева существует только некоторое допустимое подмножество возможных продолжений, которое должно быть выделено. Все это потребовало разработки оригинальных методов вычисления оценок, стратегий ветвления, а также (на основе "оптимистических" оценок возможности выполнения ограничений) разработать правила отсева недопустимых продолжений. Приведенные алгоритмы решения иллюстрируются числовыми примерами. Анализируются результаты вычислительных экспериментов.

1. Постановка задачи

Рассматриваются следующие задачи транспортной логистики.

Задача 1. Транспортное средство (ТС) должно доставить груз от нескольких производителей и складов дистрибьютеров некоторому подмножеству потребителей. Общее число пунктов загрузки и разгрузки ТС равно $n - \tilde{I} = \{i = 0, 1, \dots, n\}$. Грузоподъемность ТС ограничена и равна G . ТС в начальный момент времени находится на базе ($i = 0$), на которую должно вернуться по окончании всех перевозок. Пусть задана матрица объемов грузов, которые необходимо доставить или забрать в различных пунктах $Q = \|q_i\|$, $i = 0, 1, \dots, n$. Причем, если груз необходимо доставить в пункт i , то значение q_i задано со знаком "+", а если забрать из пункта i , то — со знаком "-". Могут возникнуть ситуации, когда для некоторых пунктов $i \in \tilde{I}$ значения $|q_i| > G$. Тогда в этот пункт придется заезжать несколько раз. Заданы матрицы размерностью $(n + 1 - C = \|c_{ij}\|$ и $A = \|a_{ij}\|$, $i, j = 0, 1, \dots, n$, соответственно стоимости и времени переезда ТС между пунктами. При этом временем и стоимостью разгрузки и погрузки будем пренебрегать либо считать не зависящими от объемов и вида перевозимого груза и учтенных в значениях c_{ij} и a_{ij} . Кроме того заданы промежутки времени $H_i \in [h_i^1, h_i^2]$, в течение которых груз необходимо доставить или забрать из пункта i . При этом могут возникнуть ситуации, когда для выполнения заданных сроков доставки грузов ТС должно некоторое время простаивать в пути или в некотором пункте. Пусть стоимость единицы времени простоя равна λ . Эти затраты также необходимо учитывать в критерии оптимальности задачи. Надо определить замкнутый маршрут движения ТС, начинающийся

и заканчивающийся в начальном пункте $i = 0$, обеспечивающий доставку и перевозку всех видов грузов в установленные заданиями сроки, а также минимальные затраты на перевозки.

Рассматриваемые ниже задачи являются частным случаем сформулированной задачи 1.

Задача 2. Ограничения на сроки обслуживания пунктов отсутствуют.

Задача 3. Объемы груза $\|q_i\|$, которые необходимо забрать или доставить в каждый из пунктов, не превышают грузоподъемности ТС, т. е. $|q_i| \leq G$, $i = 0, 1, \dots, n$.

Задача 4. Отправление груза осуществляется только из одного начального пункта $i = 0$, а во все остальные пункты $i = 1, \dots, n$ должен быть доставлен груз.

Аналогично [2—7] могут быть предложены математические модели рассматриваемых задач в виде моделей линейного и нелинейного булевого программирования большого размера, алгоритмы решения которых не учитывают свойства допустимых и оптимальных решений и требуют больших объемов вычислений.

2. Свойства допустимых и оптимальных маршрутов

В следующих разделах работы рассматриваются математические свойства и алгоритмы решения частного случая задач 1—3 в условиях, когда ТС ограниченной грузоподъемности G должно осуществить доставку и прием совместимых или одного вида грузов (соответственно в количествах q_i^+ и q_i^-) во множество пунктов $\tilde{I} = \{i = 0, 1, \dots, n\}$.

Заданы допустимые времена начала и завершения обслуживания каждого из пунктов $H_i \in [h_i^1, h_i^2]$, $i \in \tilde{I}$, а также матрицы времен $A = \|a_{ij}\|$ и стоимостных затрат $C = \|c_{ij}\|$, связанных с переездом из пункта i в пункт j , а также единичные затраты λ , связанные с простоем ТС в ожидании допустимого времени обслуживания. Необходимо построить маршрут движения ТС, начинающийся и завершающийся в пункте $i = 0$, удовлетворяющий всем ограничениям на времена обслуживания пунктов и не допускающий наличия груза на ТС в любой точке маршрута, большего значения G , который минимизировал бы суммарные затраты на выполнение расписаний.

Пусть маршрут движения ТС имеет вид $\tilde{W} = \{w_0, w_1, \dots, w_{\vartheta}, \dots, w_{\Lambda}, w_0\}$, где $\Lambda \geq n$, (так как в некоторые пункты ТС может заходить несколько раз) — суммарное число пунктов (не включая пункт $i = 0$), куда заходило ТС, и $\bar{\tau}$ — суммарное время простоев ТС в ожидании времени обслуживания. Обозначим $(w_{\vartheta-1}, w_{\vartheta}) \in \tilde{W}$, $\vartheta = 1, \dots, \Lambda$, — путь между этими двумя пунктами. Тогда функция цели задачи имеет вид

$$F(\tilde{W}) = c_{(w_0, w_1)} + \sum_{(w_{\vartheta-1}, w_{\vartheta}) \in \tilde{W}} c_{(w_{\vartheta-1}, w_{\vartheta})} + c_{(w_{\Lambda}, w_0)} + \lambda \bar{\tau}.$$

Обозначим $T(w_i^1 | \tilde{W})$, $T(w_i^2 | \tilde{W})$ — соответственно время первого и последнего посещения i -го пункта маршрута \tilde{W} . Тогда ограничения на выполнение сроков обслуживания пунктов могут быть представлены в виде

$$T(w_i^1 | \tilde{W}) \geq h_i^1; T(w_i^2 | \tilde{W}) \leq h_i^2.$$

Ниже исследуются свойства допустимых и оптимальных маршрутов и алгоритмы решения этих задач.

Назовем $\tilde{U}^s = \{0, i_1^s, i_2^s, \dots, i_l^s, \dots, i_p^s\}$ — некоторый частичный маршрут движения ТС; $\tilde{J}^{s,+}$ и $\hat{J}^{s,-}$ — соответственно подмножество пунктов, для которых выполнены все задания и которые необходимо еще посетить, реализуя частичный маршрут (план) \tilde{U}^s , включающий подмножество пунктов $i_l^s \in \tilde{J}^{s,+}$; $\hat{J}^{s,-}$, куда необходимо зайти на s -м шаге вычислений; l^s — последний пункт частичного плана \tilde{U}^s ; $\theta(\tilde{U}^s)$ — время завершения всех заданий в конечном пункте l^s в маршруте \tilde{U}^s , который должен быть продолжен дальше из этого пункта; $|q_j^s|$ — абсолютное значение объема грузов, которое еще необходимо доставить или забрать из пунктов $j \in \hat{J}^{s,-}$, развивая маршрут \tilde{U}^s ; G^s — объем груза, содержащегося в ТС при выезде из пункта l^s в маршруте \tilde{U}^s ; G — грузоподъемность транспортного средства. Обозначим соответственно $\tilde{I}^{s,+}$ и $\tilde{I}^{s,-}$ — подмножество пунктов, куда необходимо доставить и откуда забрать груз, и наличие груза в ТС и(или) возможность дозагрузки его предоставляют такие возможности,

$$\tilde{I}^{s,+} = \{i \in \hat{J}^{s,-} | q_i^s > 0\}, \tilde{I}^{s,-} = \{i \in \hat{J}^{s,-} | q_i^s < 0\}. (1)$$

В начале итеративного процесса $\hat{J}^{0,-} = \tilde{I} = \{0, 1, \dots, n\}$. Движение, как и в задаче 1, начинается и должно завершиться в начальном пункте $i = 0$. Определим

$$\omega_i = \min_{j \in \hat{J}^{s,-}} a_{ij}, \vartheta_{ij} = a_{ij} - \omega_j, i \in \hat{J}^{s,-}; (2)$$

$$\beta_j = \min_{i \in \hat{J}^{s,-}} \vartheta_{ij}, i \in \hat{J}^{s,-}; (3)$$

$$\bar{a}_j = \omega_j + \beta_j, i \in \hat{J}^{s,-}. (4)$$

В каждой строке и каждом столбце преобразованной матрицы $\hat{A} = \|\hat{a}_{ij}\|$, где $\hat{a}_{ij} = \vartheta_{ij} - \beta_j$, $i, j \in \hat{J}^{s,-}$, содержится по крайней мере по одному нулевому элементу.

Утверждение 1. Если найдется хотя бы один пункт $j \in \hat{J}^{s,-}$, для которого справедливо неравенство

$$\theta(\tilde{U}^s) + \bar{a}_j > \min_{1 \leq p \leq P} \{h_{jp}^2 | |q_{jp}^s| > 0\}, j \in \hat{J}^{s,-}, (5)$$

где P — число различных видов грузов (в данном случае $P = 1$), то развиваемый маршрут \tilde{U}^s не содержит допустимых маршрутов и может быть отброшен как неперспективный.

Упорядочим все пункты подмножества $j \in \hat{J}^{s,-}$ в последовательность

$$\tilde{W}^s = \{j_1^s, j_2^s, \dots, j_r^s, \dots, j_R^s | h_{j_r}^{2,s} \leq h_{j_{r+1}}^{2,s}; j_r^s \in \hat{J}^{s,-}\}. (6)$$

Утверждение 2. Если найдется хотя бы один пункт $j \in \hat{J}^{s,-}$, для которого не выполняется хотя бы одно из системы неравенств

$$\theta(\tilde{U}^s) + \sum_{\rho=1}^r \bar{a}_{j_\rho} \leq h_{j_r}^{2,s}, j_r \in \tilde{W}^s, r = 1, 2, \dots, R, (7)$$

где $h_{j_r}^{2,s}$ — допустимое наиболее позднее время доставки или приема грузов в пункте j_r ; r и ρ — бегущие индексы, R — число подлежащих обслуживанию пунктов в подмножестве $\hat{J}^{s,-}$, то развиваемый маршрут \tilde{U}^s не содержит допустимых планов движения и может быть отброшен как неперспективный.

Доказательства утверждения 2 приведены в книге автора [3].

3. Обслуживание пунктов

При продолжении некоторого частичного маршрута \tilde{U}^m в маршрут \tilde{U}^s с конечным пунктом $k = l^s$, вектор значений объемов грузов во всех пунктах q_j^s , $j \in \hat{J}^{s,-}$, и объем груза, оставшегося в ТС, определяются по формулам

$$q_j^s = \begin{cases} q_j^m, & \text{если } j \neq k, j \in \hat{J}^{s,-}, \\ q_k^m - \min(G^s, q_k^m), & \text{если } q_k^m > 0, \\ q_k^m + \min(G - G^s, |q_k^m|), & \text{если } q_k^m < 0; \end{cases} (8)$$

$$G^s = \begin{cases} G^m - (q_k^m - q_k^s), & \text{если } q_k^m < 0. \\ G - G^m + (|q_k^m| - |q_k^s|), & \text{если } q_k^m < 0. \end{cases}$$

Утверждение 3. Если $G^s = 0$, а $q_{jp}^s > 0$, либо $G - G^s = 0$, а $q_j^s < 0$, то маршрут из конечного пункта l^s в пункты j является нецелесообразным и не может рассматриваться в оптимальном маршруте ТС.

Правило отсева 1.

Если для параметров двух рассматриваемых частичных маршрутов \tilde{U}^s и \tilde{U}^m с функциями цели (время или стоимость выполнения маршрута) соответственно f^s и f^m , которые заканчиваются в одном

и том же пункте $l^s = l^m$, справедлива следующая система неравенств:

$$f^s \leq f^m, \quad (9)$$

$$G^s \leq G^m \text{ и } q_j^s \leq G^s, j \in \tilde{I}^{s,+}$$

$$\text{или } G^s \geq G^m, |q_j^s| \leq G - G^s, j \in \tilde{I}^{s,-}, \quad (10)$$

и если хотя бы одно из этих неравенств выполняется как строгое неравенство, то частичный план \tilde{U}^s является более перспективным, чем частичный план \tilde{U}^m , и последний может быть отброшен из рассмотрения как неперспективный.

Преобразование 1.

Выполняем следующие преобразования матрицы $\bar{C}^s = \|\bar{c}_{ij}^s\|$:

1. Если при продолжении маршрута \tilde{U}^m с конечным пунктом l^m в пункт l^s построен новый маршрут \tilde{U}^s , для которого $q_{l^m}^m = 0$, то вычеркиваем l^m -ю строку в матрице \bar{C}^m .

Если справедливо одно из условий $q_{l^s}^m \leq G^m \forall j \in \tilde{I}^{s,+}$ или $|q_{l^s}^m| \leq G - G^m \forall j \in \tilde{I}^{s,-}$, то вычеркиваем l^s -й столбец в матрице \bar{C}^m . В данном случае элемент $\bar{c}_{l^s l^m}^m$ матрицы \bar{C}^m полагаем равным ∞ . Преобразованную матрицу обозначим \bar{C}^s .

2. Если $\theta(\tilde{U}^s) + a_{l^s j} > h_j^2$, то полагаем $\bar{c}_{l^s j}^s = \infty$.

3. Выполняем следующие вычисления для всех строк $i \in \hat{J}^{s,-}$:

$$\bar{\delta}_i^s = \min_{j \in \hat{J}^{s,-}} \bar{c}_{ij}^s, \bar{\lambda}_{ij}^s = \bar{c}_{ij}^s - \bar{\delta}_i^s, i \in \hat{J}^{s,-}. \quad (11)$$

4. Для каждого из столбцов $i \in \hat{J}^{s,-}$ преобразуем элементы матрицы следующим образом:

$$L^s = \|\bar{\lambda}_{ij}^s\|, i, j \in \hat{J}^{s,-}. \quad (12)$$

Обозначим $\bar{c}_{ij}^s = \bar{\lambda}_{ij}^s - \bar{\gamma}_j^s, j \in \hat{J}^{s,-}; \bar{\gamma}_j^s = \min_{i \in \hat{J}^{s,-}} \bar{\lambda}_{ij}^s$.

5. Вычислим сумму приведенных констант

$$D^s = \sum_{i \in \hat{J}^{s,-}} \bar{\delta}_i^s + \sum_{i \in \hat{J}^{s,-}} \bar{\gamma}_j^s. \quad (13)$$

В начале процесса полагаем $C^0 = C$ и вычисленная сумма приводящих констант равна D^0 .

Утверждение 4. ТС, развивая маршрут \tilde{U}^s , должен зайти в каждый из пунктов $j \in \hat{J}^{s,-}$ не менее, чем y_j^s раз, где

$$y_j^s = 1 + \left\lfloor \frac{q_j^s - G^s}{G} \right\rfloor \forall j \in \tilde{I}^{s,+}$$

$$\text{или } y_j^s = 1 + \left\lfloor \frac{-q_j^s - (G - G^s)}{G} \right\rfloor \forall j \in \tilde{I}^{s,-}. \quad (14)$$

Здесь $\left\lfloor \frac{q_j^s - G^s}{G} \right\rfloor$ и $\left\lfloor \frac{-q_j^s - (G - G^s)}{G} \right\rfloor$ — наибольшие

целые числа частных от деления соответствующих величин.

Утверждение 5. Нижняя граница суммарных затрат вычисляется по формуле

$$\xi(\tilde{U}^0) = \sum_{j \in \tilde{I}} y_j^0 (\bar{\delta}_j^0 + \bar{\gamma}_j^0). \quad (15)$$

Утверждение 6. Если продолжением некоторого маршрута \tilde{U}^m в пункт l^s получен частичный маршрут \tilde{U}^s , то нижняя граница суммарных затрат вычисляется по формуле

$$\xi(\tilde{U}^s) = \xi(\tilde{U}^m) + D^s + \sum_{j \in \hat{J}^{s,-}} (\bar{\delta}_j^0 + \bar{\gamma}_j^0) \cdot \max[0; (y_j^s - y_j^m)] + \lambda \tau^s, \quad (16)$$

где $\tau^s = \max(0, h_1^s \theta(U^s))$.

4. Алгоритм решения задачи методом динамического программирования и последовательного анализа вариантов

В процессе решения задачи выполняется построение, анализ различных допустимых и конкурирующих маршрутов \tilde{U}^s движения ТС, начиная с начального пункта $i = 0$, и развитие перспективных маршрутов до получения наиболее эффективного решения. В процессе построения проводится отсев недопустимых и неперспективных продолжений. Каждая итерация алгоритма ($s = 0, 1, 2, \dots, S$) состоит из нескольких шагов. Развивается перспективный s^* -й план с наименьшим значением функции цели, ($f^{s^*} = \min_{s=1, \dots, S} f^s$), который продолжается во все допустимые пункты его продолжения. Проводится отсев всех неперспективных частичных планов. Первый допустимый план перевозок, значение критерия оптимальности которого меньше или равно значению суммарных затрат любого из конкурирующих и не продолженных до конца частичных планов, является решением задачи. На основе установленных выше свойств допустимых решений задачи и системы оценок нижних границ значений функции цели на ранних стадиях построения альтернативных маршрутов движения анализируется возможность выполнения системы ограничений и проводится отсев недопустимых продолжений.

Приведенные выше правила сравнения альтернативных частичных планов позволяют отсеять строящиеся маршруты, не содержащие оптимальных решений. Тот факт, что на каждой итерации развивается только частичный план (строящийся маршрут) с наилучшим значением функции цели, позволяет исключить анализ заведомо неоптимальных маршрутов.

Метка каждого развивающегося частичного плана \tilde{U}^s включает следующие параметры:

$$\tilde{U}^s = \{0, i_1^s, i_2^s, \dots, i_\lambda^s, \dots, i_l^s\}; l^s, \theta(\tilde{U}^s); \bar{J}^{s,+} \text{ и } \hat{J}^{s,-};$$

$$q_{ip}^s, i = 0, 1, \dots, n; p \in \tilde{P} = \{1, \dots, P\}; G^{s,+} = \|g_p^{s,+}\|,$$

где $g_p^{s,+}$ — количество p -го вида груза, находящегося в ТС; $f^s = \infty$; $F(Z^s)$, где Z^s — построенный допустимый маршрут движения ТС, а $F(Z^s)$ — значение функции цели для построенного допустимого маршрута; $\psi(\tilde{U}^s)$ — параметр, определяющий, продолжен ли этот маршрут во все допустимые пункты: если да, то $\psi(\tilde{U}^s) = 1$, если нет, то $\psi(\tilde{U}^s) = 0$, при $\psi(\tilde{U}^s) = 2$ маршрут исключен из дальнейшего рассмотрения как недопустимый, а при $\psi(\tilde{U}^s) = 3$ — как неоптимальный.

Пусть на некотором этапе решения задачи получено $Z^t, t = 1, \dots, T$, допустимых маршрутов решения задачи, функции цели которых соответственно равны $F(Z^t)$, среди которых $F(\bar{Z}^T) = \min_{1 \leq t \leq T} F(Z^t)$ — наилучшее среди всех полученных на данный момент решений задачи.

В начале процесса решения ($s = 0$) формируем для частичного маршрута \tilde{U}^0 следующие параметры начальной метки:

$$\tilde{U}^0 = \{0\}; l^0 = 0; \theta(\tilde{U}^0) = 0; \bar{J}^{0,+} = \tilde{I} = \{0, 1, 2, \dots, n\},$$

$$\hat{J}^{0,-} = \emptyset; q_i^0 = q_i, i = 0, 1, \dots, n; G^0 = 0; f^0 = 0;$$

$$F(\bar{Z}^T) = \infty; \psi(\tilde{U}^0) = 0.$$

На каждой итерации алгоритма выполняем следующий объем вычислений.

Шаг 1. Среди всех частичных планов ($s = 0, 1, 2, \dots, S$) выбираем некоторый частичный план \tilde{U}^n , для которого $f^n = \min_{0 \leq s \leq S} \{f^s | \psi(\tilde{U}^s) = 0\}$. Переходим к шагу 2.

Шаг 2. Рассматриваем частичный план \tilde{U}^n , конечный пункт маршрута которого l^n . Из пунктов $j \in \hat{J}^{n,-}$ находим такие, для которых выполняется неравенство

$$\theta(\tilde{U}^n) + a_{l^n j} \leq \begin{cases} h_j^2, & \text{если } q_j^n > 0 \text{ \& } G^n > 0, \\ h_j^2, & \text{если } q_j^n < 0 \text{ \& } (G - G^n) > 0. \end{cases} \quad (17)$$

Пусть это будет подмножество пунктов $j \in \hat{J}_1^{n,-} \subseteq \hat{J}^{n,-}$.

А. Если $\hat{J}_1^{n,-} = \emptyset$, т. е. таких пунктов не существует, то частичный план \tilde{U}^n не содержит допустимых решений, и он исключается из дальнейшего рассмотрения. В этом случае полагаем $\psi(\tilde{U}^n) = 2$, выполняем переиндексацию всех рассматриваемых частичных планов и вновь переходим к шагу 1.

Б. Если $\hat{J}_1^{n,-} \neq \emptyset$, то для каждого из пунктов $j \in (\hat{J}_1^{n,-} / l^n)$ определяем

$$\theta(j | \tilde{U}^n) = \max \{ h_j^1; \theta(\tilde{U}^n) + a_{l^n j} \}, j \in (\hat{J}_1^{n,-} / l^n). \quad (18)$$

$$\tilde{V}^n = \{j_1^n, j_2^n, \dots, j_{k-1}^n, j_k^n, \dots, j_{K(n)}^n | j_k^n \in (\hat{J}_1^{n,-} / l^n);$$

$$h_{j_{k-1}}^2 \leq h_{j_k}^2, k = 2, \dots, K\}. \quad (19)$$

Проверяем выполнение следующей системы неравенств для индексов, упорядоченных в последовательность (19):

$$\theta(\mathcal{G} | \tilde{U}^n) + a_{l^n \mathcal{G}} + \sum_{\delta=1}^k \bar{a}_{j_\delta^n} - \bar{a}_{\mathcal{G}} \leq h_{j_k^n}^2, i_k^n \in (\tilde{V}^n / \mathcal{G}). \quad (20)$$

Если для некоторого пункта j_k^n не выполняется хотя бы одно из системы неравенств (20), то в допустимых продолжениях частичного плана \tilde{U}^n маршрут из пункта l^n в \mathcal{G} является недопустимым.

В. Обозначим \tilde{M}^n подмножество пунктов $\mathcal{G} \in (\hat{J}_1^{n,-} / l^n)$, для которых справедливы все неравенства (18)—(20). Если таких пунктов не существует, т. е. $\tilde{M}^n = \emptyset$, то \tilde{U}^n исключается из дальнейшего рассмотрения, полагаем $\psi(\tilde{U}^n) = 1$, проводим переиндексацию всех рассматриваемых частичных планов и переходим к шагу 1. Если $\tilde{M}^n \neq \emptyset$, то переходим к шагу 3.

Шаг 3. Продолжаем частичный план \tilde{U}^n во все пункты подмножества $\xi_r \in \tilde{M}^n$.

При этом метка каждого из продолженных путей формируется следующим образом:

$$\bar{V}^{S+r} = \{0, i_1^n, i_2^n, \dots, i_\lambda^n, \dots, i_l^n, i_{\xi_r}^{S+r}\}; l^{S+r};$$

$$\bar{J}^{(S+r),+} = (\bar{J}^{n,+} \cup \xi_r); \bar{J}^{(S+r),-} = (\bar{J}^{n,-} / \xi_r);$$

$$\theta(\tilde{U}^{S+r}) =$$

$$= \begin{cases} \theta(\tilde{U}^n) + a_{l^n \xi_r}, & \text{если } \theta(\tilde{U}^n) + a_{l^n \xi_r} \geq h_{\xi_r}^1, \\ h_{\xi_r}^1, & \text{если } \theta(\tilde{U}^n) + a_{l^n \xi_r} < h_{\xi_r}^1; \end{cases} \quad (21)$$

$$q_i^{(S+r)} = q_i^n, i \in (\bar{J}^{n,-} / \xi_r).$$

Значения $q_{\xi_r}^{(S+r)}$ и $G^{(S+r)}$ вычисляются по формулам (8):

$$q_{\xi_r}^{(S+r)} = \begin{cases} q_{\xi_r}^n - \min(G^n, q_{\xi_r}^n), & \text{если } q_{\xi_r}^n > 0, \\ q_{\xi_r}^n + \max(G - G^n, |q_{\xi_r}^n|), & \text{если } q_{\xi_r}^n < 0; \end{cases} \quad (22)$$

$$G^{(S+r)} = \begin{cases} G^n - (q_k^{(S+r)} - q_{\xi_r}^s), & \text{если } q_k^n < 0, \\ G - G^n + (|q_{\xi_r}^{(S+r)}| - |q_{\xi_r}^s|), & \text{если } q_k^n < 0. \end{cases} \quad (23)$$

$$f(\tilde{U}^{(S+r)}) =$$

$$= \begin{cases} f(\tilde{U}^n) + c_{l_{\eta_{\xi_r}}}, & \text{если } h_{\xi_r}^1 \leq (\theta(\tilde{U}^n) + a_{l_{\eta_{\xi_r}}}), \\ f(\tilde{U}^n) + c_{l_{\eta_{\xi_r}}} + \lambda[h_{\xi_r}^1 - \theta(\tilde{U}^n) - a_{l_{\eta_{\xi_r}}}], & \\ \text{если } h_{\xi_r}^1 > (\theta(\tilde{U}^n) + a_{l_{\eta_{\xi_r}}}). \end{cases} \quad (24)$$

Полагаем $\psi(\tilde{U}^n) = 1$, $\psi(\tilde{U}^{(S+r)}) = 0$, $r = 1, \dots, R$. Переходим к шагу 4.

Шаг 4. Исключаем из дальнейшего рассмотрения все продолженные на этой итерации маршруты $\tilde{U}^{(S+r)}$, для которых справедливо неравенство $F(\tilde{U}^{(S+r)}) \geq F(\bar{Z}^T)$, положив для них $\psi(\tilde{U}^{(S+r)}) = 3$.

А. Если для какого-то вновь образованного маршрута $\tilde{U}^{(S+r)}$ выполняются условия $\hat{J}_1^{(S+r),-} = \emptyset$, $l^{(S+r)} = 0$; $q_j^{(S+r)} = 0, j \in \tilde{I}$; $G^{(S+r)} = 0$, то получено допустимое решение задачи, значение функции цели которого (время или стоимость выполнения маршрута) равно $F(\tilde{U}^{(S+r)})$. Определяем для этого маршрута $\psi(\tilde{U}^{(S+r)}) = 1$, $T := (T + 1)$.

Если $F(\tilde{U}^{(S+r)}) < F(\bar{Z}^T)$, то полагаем $Z^T = \tilde{U}^{(S+r)}$, $F(\bar{Z}^T) = F(\tilde{U}^{(S+r)})$.

Если выполняются условия $F(\bar{Z}^T) \leq \{f(\tilde{U}^\tau) | \psi(\tilde{U}^\tau) = 0\}$ для всех $\tau = 0, 1, 2, \dots, S, (S+1), \dots, (S+R)$, то этот маршрут со значением функции цели, равным $F(\bar{Z}^T)$, является решением задачи, и алгоритм завершает работу.

Б. Для каждого из вновь образованных маршрутов, положив $\tilde{U}^m = \tilde{U}^{(S+r)}$ проверяем выполнение условий правила отсева 1, сравнивая каждый из них со всеми частичными планами \tilde{U}^s , у которых признак $\psi(\tilde{U}^s) = 0$. Если выполняются все условия правила отсева 1, то $\tilde{U}^m = \tilde{U}^{(S+r)}$ является неперспективным, полагаем $\psi(\tilde{U}^{(S+r)}) = 3$, исключаем его из дальнейшего рассмотрения. Если для некоторой

пары индексов s и $(S+r)$ условия правила отсева 1 выполняются в условиях обозначений $\tilde{U}^s = \tilde{U}^{(S+r)}$, а $\tilde{U}^m = \tilde{U}^s$, ранее построенный частичный план \tilde{U}^s является неперспективным, полагаем $\psi(\tilde{U}^s) = 3$, и исключаем его рассмотрения.

Переходим к следующей итерации, возвратившись к шагу 1.

Через некоторое число итераций алгоритма либо будет оптимальное решение задачи, либо будет установлен факт, что исходная система ограничений задачи является несовместной.

4. Алгоритм решения задачи модифицированным методом ветвей и границ

Метка каждого развивающегося частичного плана \tilde{U}^s включает следующие параметры: $\tilde{U}^s = \{0, i_1^s, i_2^s, \dots, i_\lambda^s, \dots, i_l^s\}; l^s, \theta(\tilde{U}^s); \bar{J}^{s,+}$ и $\bar{J}^{s,-}; q_i^s, i = 0, 1, \dots, n; G^s; f^s; \psi(\tilde{U}^s)$ — параметр, определяющий, продолжен ли этот маршрут во все допустимые пункты: если да — $\psi(\tilde{U}^s) = 1$, если нет — $\psi(\tilde{U}^s) = 0$. При $\psi(\tilde{U}^s) = 2$ маршрут недопустимый, а при $\psi(\tilde{U}^s) = 3$ — неоптимальный.

На начальном этапе решения ($s = 0$), когда $G^0 = 0$, выполним следующие преобразования и вычисления:

- в 0-й строке матриц C^0 и A^0 полагаем значения $c_{0j} = a_{0j} = \infty$ в тех столбцах этой строки, для которых справедливы соотношения $q_j < 0$, а $q_0 = G$;
- рассмотрим последовательность пунктов

$$\tilde{W}^0 = \{i_1, i_2, \dots, i_r, \dots, i_r | h_{i_r}^2 \leq h_{i_{r+1}}^2; i_r, i_{r+1} \in \tilde{I}\}; \quad (25)$$

проверяем выполнение следующей системы неравенств:

$$\sum_{p=1}^r \bar{a}_{i_p} > h_{i_r}^2, i_r \in \tilde{W}^0, r = 1, 2, \dots, R; \quad (26)$$

если справедливо неравенство (26) хотя бы для одного индекса $i_r \in \tilde{W}^0, r = 1, 2, \dots, R$, то система ограничений задачи является несовместной, и алгоритм завершает свою работу;

в) в противном случае вычислим значения $\bar{a}_i, i \in \tilde{I}$, по формулам (3)—(5), а также значения $\bar{\delta}_j^0, \bar{\gamma}_j^0, y_j^0$ соответственно по формулам (11), (12), (14) и значение нижней границы функции цели $\xi(\tilde{U}^0)$ по формуле (15);

г) положив $s = 0, C^0 = C = \|c_{ij}\|$, выполним преобразование 1 матрицы C^0 . Преобразованную по пра-

вину 1 матрицу $C^s, s = 0, 1, \dots, S$, в дальнейшем будем обозначать $\bar{C}^s = \|\bar{c}_{ij}^s\|, i, j \in \hat{J}^{s,-}$; полагаем $F(\bar{Z}^0) = \infty$.

Сформируем метку начальной вершины дерева различных вариантов:

$$\tilde{U}^{s=0} = \{0\}, l^0 = 0; \theta(\tilde{U}^0) = 0; \bar{J}^{s,+} = \{0\}, \hat{J}^{s,-} = \tilde{I},$$

где \tilde{I} — множество пунктов, подлежащих обслуживанию в соответствии с условиями задачи;

$$f^0 = F(\bar{Z}^0) = \infty;$$

$$q_0^0 = \begin{cases} q_0, & \text{если } q_0 < 0 \text{ \& } G^0 = 0, \\ q_0 - \min(q_0, G^0), & \text{если } q_0 < 0 \text{ \& } G^0 > 0; \end{cases}$$

$$G^0 = G - q_0^0; \xi(Z^0); \psi(\tilde{U}^0) = 0.$$

Пусть на некотором этапе решения получено $s = 1, \dots, S$ различных частичных и требующих дальнейшего развития маршрутов движения ТС (вершин дерева рассматриваемых вариантов) \tilde{U}^s , вид метки которых приведен выше. Кроме того, получено некоторое подмножество допустимых решений (альтернативных маршрутов), значение суммарных приведенных затрат для лучшего из которых равно $F(\bar{Z}^s)$. Алгоритм решения задачи предусматривает выполнение следующих шагов.

Шаг 1. Для всех вершин (меток) дерева, для которых $\psi(\tilde{U}^s) = 0$, если $\xi(Z^s) \geq F(\bar{Z}^s)$, полагаем $\psi(\tilde{U}^s) = 3$. Находим метку с индексом $\bar{s} = \arg\{\min_{0 \leq s \leq S} \xi(Z^s) | \psi(\tilde{U}^s) = 0\}$.

Определяем для этой вершины дерева подмножества $\bar{J}^{\bar{s},+}, \bar{J}^{\bar{s},-}$, пункт $k = l^{\bar{s}}$, матрицы $\bar{C}^{\bar{s}} = \|\bar{c}_{ij}^{\bar{s}}\|, i, j \in \hat{J}^{\bar{s},-}$, а также значение $\bar{G}^{\bar{s}}$ и вектор $\bar{q}_j^{\bar{s}}, j \in \hat{J}^{\bar{s},-}$. В k -й строке матриц $C^{\bar{s}}$ и $A^{\bar{s}}$ полагаем значения $c_{kj}^{\bar{s}} = a_{kj}^{\bar{s}} = \infty$ в тех столбцах этой k -й строки, для которых справедливы соотношения $q_j^{\bar{s}} \leq 0$, а $G - \bar{G}^{\bar{s}} = 0$. Переходим к шагу 2.

Шаг 2. Вычисляем значения суммы приводящих констант для каждого пункта $\bar{a}_k^{\bar{s}} = \min_{j \in \hat{J}^{\bar{s},-}} \bar{a}_{kj}^{\bar{s}}, k \in \hat{J}^{\bar{s},-}$. Упорядочим все пункты подмножества $j \in \hat{J}^{\bar{s},-}$ в последовательность $\bar{W}^{\bar{s}}$ в соответствии с выражением (6). Проверяем выполнение системы неравенств (7). Если не выполняется хотя бы одно из неравенств (7), то этот маршрут не содержит допустимых продолжений, полагаем $\psi(\tilde{U}^{\bar{s}}) = 2$, и переходим к шагу 1. В противном случае выполняем

преобразование 1 матрицы $\bar{C}^{\bar{s}} = \|\bar{c}_{ij}^{\bar{s}}\|, i, j \in \hat{J}^{\bar{s},-}$, вычисляем сумму приводящих констант и переходим к шагу 3.

Шаг 3. Продолжаем путь $\tilde{U}^{\bar{s}}$ (развиваем \bar{s} -ю вершину дерева), продолжив путь из вершины $k = l^{\bar{s}}$ во все вершины с индексом, значение соответствующего элемента которого в этой k -й строке равно 0 — $\{j_1 \in \hat{J}^{\bar{s},-} | \bar{c}_{kj_1}^{\bar{s}} \neq \infty\}$.

Для вновь образованных η вершин $\rho = (S + 1), (S + 2), \dots, (S + \eta)$, метка формируется следующим образом:

$$\tilde{U}^\rho = \{0, i_1^{\bar{s}}, i_2^{\bar{s}}, \dots, i_\lambda^{\bar{s}}, \dots, l_{\bar{s}} = k, j_1(\rho)\}, l^\rho = j_1(\rho), (27)$$

$$\theta(\tilde{U}^\rho) =$$

$$= \begin{cases} \theta(\tilde{U}^s) + a_{k,j_1(\rho)}, & \text{если } \theta(\tilde{U}^s) + a_{k,j_1(\rho)} \geq h_{j_1(\rho)}^1, \\ h_{j_1(\rho)}^1, & \text{если } \theta(\tilde{U}^s) + a_{k,j_1(\rho)} < h_{j_1(\rho)}^1; \end{cases} (28)$$

$$\bar{J}^{\rho,+} = \bar{J}^{\bar{s},+} \cup j_1(\rho); \hat{J}^{\rho,-} = \{\hat{J}^{\bar{s},-} / j_1(\rho)\}. (29)$$

Значения $q_i^\rho, i \in \hat{J}^{\rho,-}$ и G^ρ определяются в соответствии с алгоритмом обслуживания пунктов по формулам (22), (23).

Другие показатели метки для вновь образованных частичных маршрутов определяются по формулам

$$\xi(Z^\rho) = \xi(Z^{\bar{s}}) + \sum_{j \in \hat{J}^{\bar{s},-}} d_j \cdot \psi_j^{\bar{s}} + \bar{c}_{k,j_1(\rho)}^{\bar{s}} + b \cdot \max[0; h_j^1 - \theta(\tilde{U}^{\bar{s}}) - a_{k,j_1(\rho)}]; (30)$$

$$f^\rho = \begin{cases} \xi(Z^\rho), & \text{если } \hat{J}^{\rho,-} = \emptyset, \\ \infty, & \text{если } \hat{J}^{\rho,-} \neq \emptyset; \end{cases} \psi(\tilde{U}^\rho) = \begin{cases} 1, & \text{если } f^\rho < \infty, \\ 0, & \text{если } f^\rho = \infty. \end{cases} (31)$$

Для \bar{s} -го частичного маршрута полагаем $\psi(\tilde{U}^{\bar{s}}) = 1$. Переходим к шагу 4.

Шаг 4. Для вновь образованных вершин $\rho = (S + 1), (S + 2), \dots, (S + \eta)$ определяем $\bar{f} = \min_{(S+1) \leq \rho \leq (S+\eta)} f^\rho$. Если $\bar{f} < \infty$ и $\bar{f} < F(\bar{Z}^s)$, то полагаем $\bar{f} < F(\bar{Z}^s)$. Если $F(\bar{Z}^s) \leq \min_{1 \leq s \leq (S+\eta)} \xi(Z^s)$, т. е.

значение целевой функции полученного допустимого решения меньше или равно, чем нижняя граница функции цели всех висячих и подлежащих развитию вершин дерева, то получено решение задачи и алгоритм завершает работу. В противном случае переходим к шагу 1.

Через некоторое число больших итераций алгоритма либо будет получено оптимальное решение задачи, либо будет установлен факт несовместности исходной системы ограничений.

5. Иллюстративные примеры

Таблица 1

Иллюстративный пример 1.

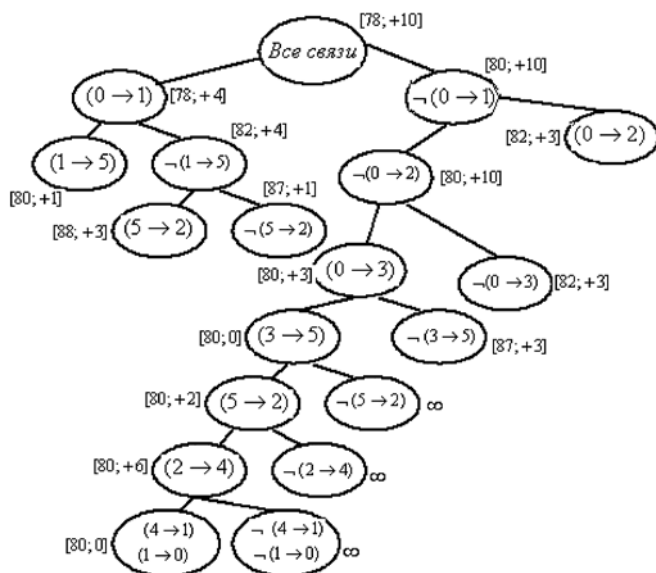
ТС должно осуществить доставку и получение грузов из шести пунктов. Грузоподъемность транспортного средства — не более 12 т. Маршрут движения должен начинаться и заканчиваться в начальном пункте $i = 0$. Расстояния (время проезда или стоимости проезда из одного пункта в другой) приведены в табл. 1. Объемы грузов, которые необходимо доставить или забрать в различных пунктах $i = 0, 1, \dots, 5$, приведены в табл. 2. Необходимо найти допустимый маршрут движения транспортного средства, обеспечивающий минимум затрат.

Дерево решений задачи представлено на рисунке. В табл. 3 дана приведенная матрица расстояний в процессе назначения и запрещения отдельных связей на главной ветви дерева в процессе решения задачи.

Следовательно, оптимальным решением задачи является маршрут движения $(0 \rightarrow 3 \rightarrow 5 \rightarrow 2 \rightarrow 4 \rightarrow 1 \rightarrow 0)$, длина (стоимость) которого составляет $14 + 15 + 10 + 17 + 13 + 11 = 80$ км (тыс.руб).

Иллюстративный пример 2.

ТС должно осуществить доставку и получение грузов из семи пунктов. Грузоподъемность транспортного средства не более 11 т. Маршрут движения должен начинаться и заканчиваться в начальном пункте $i = 0$. Время проезда или стоимости проезда из одного пункта в другой приведены в табл. 4, а стоимость перевозок — в табл. 5. Каждая единица времени простоя ТС связана с затратами, значения которых равны 1. Объемы грузов, которые необходимо доставить или забрать в различных пунктах $i = 0, 1, \dots, 6$, приведены в табл. 6 (со знаком "+" заданы объемы грузов, которые нужно забрать в различных пунктах, а со знаком "-" объемы, которые



Дерево решений задачи

Матрица расстояний между пунктами

№ пунктов	Расстояния (стоимости проезда) между пунктами					
	0	1	2	3	4	5
0	∞	10	12	14	15	20
1	11	∞	13	20	19	13
2	12	15	∞	18	17	14
3	13	18	20	∞	27	15
4	16	13	15	25	∞	20
5	17	12	10	12	22	∞

Таблица 2

Объемы грузов, которые необходимо доставить или забрать в различных пунктах

Объемы грузов (в тоннах), поставляемых и забираемых из пунктов					
0	1	2	3	4	5
+10	-6	+2	-7	+4	-3

Таблица 3

Приведенная матрица расстояний между пунктами

№ пунктов	Расстояния (стоимости проезда) между пунктами					
	0	1	2	3	4	5
0	∞	0	2	2	∞	8
1	0	∞	2	7	3	0
2	0	3	∞	4	0	0
3	0	5	7	∞	9	0
4	3	0	2	10	∞	5
5	7	2	0	0	7	∞

Таблица 4

Время проезда между пунктами

№ пунктов	Время проезда между пунктами						
	0	1	2	3	4	5	6
0	∞	20	15	25	17	27	12
1	20	∞	30	10	16	23	25
2	15	30	∞	32	18	25	16
3	25	10	32	∞	20	24	35
4	17	16	18	20	∞	18	22
5	27	23	25	24	18	∞	18
6	12	25	16	35	22	18	∞

Таблица 5

Стоимость проезда между пунктами

№ пунктов	Стоимость проезда между пунктами						
	0	1	2	3	4	5	6
0	∞	27	22	35	24	38	18
1	27	∞	42	14	23	32	34
2	22	42	∞	45	25	35	23
3	35	14	45	∞	27	32	48
4	24	23	25	27	∞	25	30
5	38	32	35	32	25	∞	24
6	18	34	23	48	30	24	∞

необходимо доставить. Граничные времена обслуживания каждого из пунктов приведены в табл. 7. Грузоподъемность ТС — 11 т. Необходимо найти допустимый маршрут движения транспортного средства, обеспечивающий обслуживание каждого из пунктов в заданные ограничениями сроки и минимум затрат, связанных с выполнением всех заданий.

Выполнив приведение 1 матриц A и C , вычислим сумму приводящих констант для каждого из индексов матрицы A , которые соответственно равны $\bar{a} = (12, 15, 17, 10, 16, 20, 12)$, а также по формуле (16) нижнюю границу функции цели в оптимальном решении, которая равна 162.

Процесс решения задачи описанным выше модифицированным алгоритмом последовательного анализа вариантов приведен ниже в приложении (табл. П.1). Номера частичных маршрутов, отмеченных курсивом, определяют оптимальное решение задачи. Результаты решения задачи сведены в табл. 8.

Оптимальным решением задачи является маршрут движения $(0 \rightarrow 2 \rightarrow 5 \rightarrow 3 \rightarrow 1 \rightarrow 4 \rightarrow 6 \rightarrow 4 \rightarrow 0)$, значение критерия оптимальности для которого равно 220.

6. Вычислительные эксперименты

Автором были проведены вычислительные эксперименты анализа эффективности предлагаемых алгоритмов. Решались задачи, включающие до 20 пунктов. Следует заметить, что при решении практических задач построения маршрута движения одной машины редко можно встретить задачи большей размерности. Выбирались достаточно жесткие ограничения на сроки доставки грузов и обслуживания объектов. Время решения всех задач на персональном компьютере с процессором Core i-5 Intel® Core™ i5-4200U 4-й модификации (до 2,60 ГГц было не более 2,5 мин). Наибольшее число анализируемых вариантов и соответственно наибольший объем вычислений было в задачах без ограничений с небольшими отличиями во временах и стоимости перевозки грузов из одного пункта в другой. Наличие жесткой системы ограничений и большие разбросы во временах и стоимости доставки грузов позволяют уже на первых итерациях алгоритма исключить маршруты, не содержащие допустимых и оптимальных планов, и существенно сокращают число анализируемых маршрутов. В задачах большой размерности при несущественных ограничениях на сроки доставки грузов и небольших отличиях в значениях элементов матриц C и A объемы вычислений увеличиваются. Здесь алгоритмы метода ветвей и границ, как правило, являются более эффективными, так как происходит интенсивный отсев подмножеств маршрутов, не удовлетворяющих всей системе ограничений. При этом существенно сокращается число вершин и ветвей в дереве решений задачи и, следовательно, объем вычислений. Могут быть использованы приближенные методы решения задачи, использующие приближенные правила отсева неэффективных частичных планов. Кроме того, некоторый определенный на первых итерациях допустимый и эффективный маршрут может быть доведен до конца, не обращая внимание, что существуют маршруты, включающие меньшее число пунктов с лучшим значением критерия эффективности.

Таблица 6

Объемы грузов, которые необходимо доставить или забрать в различных пунктах

Объемы грузов (в тоннах), поставляемых и забираемых из пунктов						
0	1	2	3	4	5	6
8	-3	-6	5	-12	3	5

Таблица 7

Граничные времена обслуживания пунктов

	Граничные времена обслуживания пунктов						
	0	1	2	3	4	5	6
h_i^1	0	40	20	50	0	10	60
h_i^2	300	85	70	250	150	60	135

Таблица 8

Результаты решения иллюстративного примера 2

№ пунктов $i = l^s$	Параметры частичных маршрутов \tilde{U}^s				
	s	$\theta(\tilde{U}^s)$	q_i^s	G^s	$f^s = f(\tilde{U}^s)$
0	0	0	0	8	0
2	2	20	0	2	22
5	9	45	0	5	67
3	21	69	0	10	99
1	45	79	0	7	113
4	66	95	-5	0	136
6	72	117	0	5	166
4	77	139	0	0	196
0	81	156	0	0	220

Заключение

Рассмотрены задачи построения оптимальных маршрутов движения грузовых транспортных средств (ТС) в условиях наличия ограничений на сроки доставки и загрузки груза и на грузоподъемность ТС. В качестве критериев оптимальности рассмотрены суммарные затраты времени или в стоимостном выражении на выполнение графика доставки грузов.

Исследованы свойства допустимых и оптимальных маршрутов, на основе которых сформулированы правила отсева недопустимых и неоптимальных маршрутов уже на ранних этапах решения задачи, а также определены оценки нижней границы оп-

тимального решения для любого из строящихся в процессе решения маршрута.

На основе установленных свойств допустимых и оптимальных маршрутов предложены алгоритмы решения задачи модифицированными методами динамического программирования и последовательного анализа вариантов, которые иллюстрированы числовыми примерами.

Приведенные вычислительные эксперименты показали, что наличие жесткой системы ограничений на времена обслуживания пунктов зачастую позволяет существенно сократить число конкурирующих альтернативных маршрутов и уже на ранних этапах решения исключить недопустимые и неоптимальные варианты, а также установить несовместность сформулированной системы ограничений.

В отличие от известных методов "ветвей и границ", применяемых для решения широкого класса задач, в предлагаемых алгоритмах для висячих вершин дерева проводится также отсев неперспективных продолжений на основе сформулированных правил доминирования. Это позволяет осуществить отсев некоторого числа вершин дерева, хотя нижняя граница функции цели их лучше, чем у развиваемого подмножества вариантов, что сокращает объем вычислений.

Оригинальность предлагаемого метода динамического программирования заключается в сформулированных правилах доминирования и отсева, а также в том, что на каждом шаге развиваются не все варианты, а только тот вариант, значение функции цели которого минимально среди всех подлежащих развитию частичных планов. Это также существенно сокращает объем вычислений.

Список литературы

1. Dantzig G. B., Ramser R. H. The Truck Dispatching Problem // Management Science. 1959. N. 6. P. 80—91.
2. Зак Ю. А. Обобщенная задача коммивояжера и ее прикладные аспекты // Автоматика и телемеханика. 1976. № 10. С. 120—132.
3. Зак Ю. А. Прикладные задачи теории расписаний и маршрутизации перевозок. М.: URSS, Libroком, 2012. 393 с.
4. Бронштейн Е. М., Заико Т. А. Детерминированные оптимизационные задачи транспортной логистики // Автоматика и телемеханика. 2010. № 10. С. 133—147.
5. Vehicle Routing Problem. URL: <http://neo.lcc.uma.es/vrp/vehicle-routing-problem>. Solution Methods for VRP. URL: <http://neo.lcc.uma.es/vrp/solution-methods/>.
6. Бронштейн Е. М., Давлетбаев А. Нестационарная задача управления маршрутом транспортного средства // Проблемы управления. 2014. № 3. С. 23—28.
7. Malandraky C., Daskin M. S. Time Dependent Vehicle Routing Problems Formulations, Properties and Heuristic Algorithms // Transportation Science. 1992. Vol. 26. P. 185—200.
8. Crevier B., Cordeau J.-F., Laporte G. The multi-depot vehicle routing problem with inter-depot routes // European Journal of Operational Research. 2007. Vol. 176. P. 756—773.
9. Stegers E. A Solution Method for Vehicle Routing problems with Time-Dependent Travel Times. Delft: Delft University of Technology, 2009. 83 p.
10. Ralphs T. K., Kopman L., Pulleyblank W. R., Trotter L. E. On the Capacitated Vehicle Routing Problem // Math. Program. Ser. B. 2003. Vol. 94. P. 343—359.

Таблица П.1.1

Приведенная матрица расстояний между пунктами (запрещены связи (0 → 1) (0 → 2))

№ пунктов	Расстояния (стоимости проезда) между пунктами					
	0	1	2	3	4	5
0	∞	∞	∞	0	∞	6
1	0	∞	2	7	3	0
2	0	3	∞	4	0	0
3	0	5	7	∞	9	0
4	3	0	2	10	∞	5
5	7	2	0	0	7	∞

Таблица П.1.2

Приведенная матрица расстояний между пунктами (установлена связь (0 → 3))

№ пунктов	Расстояния (стоимости проезда) между пунктами				
	0	1	2	4	5
1	0	∞	2	3	0
2	0	3	∞	0	0
3	∞	5	7	9	0
4	3	0	2	∞	5
5	7	2	0	7	∞

Таблица П.1.3

Приведенная матрица расстояний между пунктами (установлены связи (0 → 3) и (3 → 5))

№ пунктов	Матрица расстояний между пунктами			
	0	1	2	4
1	0	∞	2	3
2	0	3	∞	0
4	3	0	2	∞
5	∞	2	0	7

Таблица П.1.4

Приведенная матрица расстояний между пунктами (установлены связи (0 → 3), (3 → 5) и (5 → 2))

№ пунктов	Матрица расстояний между пунктами		
	0	1	4
1	0	∞	3
2	∞	∞	0
4	3	0	∞

Таблица П.1.5

Приведенная матрица расстояний между пунктами (установлены связи (0 → 3), (3 → 5), (5 → 2) и (2 → 4))

№ пунктов	Матрица расстояний между пунктами	
	0	1
1	0	∞
4	∞	0

Процесс решения иллюстративного примера 2

№	\tilde{U}^s	l^s	$\theta(\tilde{U}^s)$	$\hat{J}^{s,-}$	Объемы грузов в пунктах $g_i(\tilde{U}^s)$						G^s	$f(\tilde{U}^s)$	$\psi(\tilde{U}^s)$	
					0	1	2	3	4	5				6
0	0	0	0	\tilde{I}	0	-3	-6	5	-12	3	5	8	0	1
1	(0,1)	1	40	$(\tilde{I}/1)$	0	0	-6	5	-12	3	5	5	47	2
2	(0,2)	2	20	\tilde{I}	0	-3	0	5	-12	3	5	2	32	1
3	(0,3)	3	50	$(\tilde{I}/3)$	0	-3	-6	2	-12	3	11	3	50	2
4	(0,4)	4	17	\tilde{I}	0	-3	-6	5	-4	3	5	0	24	1
5	(0,5)	5	27	$(\tilde{I}/5)$	0	-3	-6	5	-12	0	5	11	38	1
6	(0,6)	6	60	$(\tilde{I}/6)$	0	-3	-6	5	-12	3	2	11	60	2
7	(0,2,1)	1	50	$(\tilde{I}/1)$	0	-1	0	5	-12	3	5	0	50	2
8	(0,2,3)	3	52	$(\tilde{I}/3)$	0	-3	0	0	-12	3	5	7	78	2
9	(0,2,5)	5	45	$(\tilde{I}/5)$	0	-3	0	5	-12	0	5	5	67	1
10	(0,2,6)	6	36	$(\tilde{I}/1)$	0	-3	0	5	-12	3	0	7	55	1
10	(0,4,3)	3	37	$(\tilde{I}/3)$	0	-3	-6	0	-4	3	5	5	51	2
10	(0,4,5)	5	35	$(\tilde{I}/5)$	0	-3	-6	5	-4	0	5	3	49	1
11	(0,4,6)	6	60	$(\tilde{I}/6)$	0	-3	-6	5	-9	3	0	6	75	2
12	(0,5,1)	1	50	$(\tilde{I}/(1,5))$	0	0	-6	5	-12	3	5	8	56	1
13	(0,5,2)	2	52	$(\tilde{I}/(2,5))$	0	-3	0	5	-12	0	5	5	73	1
14	(0,5,4)	4	45	$(\tilde{I}/5)$	0	-3	-6	5	-1	0	5	0	63	1
16	(0,5,6)	6	46	$(\tilde{I}/(5,6))$	0	-3	-6	5	-12	0	0	0	62	2
17	(0,2,4,3)	3	58	(0,1,4,5,6)	0	-3	0	0	-10	3	5	5	84	2
18	(0,2,4,5)	5	56	(0,1,3,4,6)	0	-3	0	5	-10	0	5	3	82	2
19	(0,2,4,6)	6	60	(0,1,3,4,5)	0	-3	0	5	-10	0	0	5	87	2
20	(0,2,5,1)	1	68	(0,3,4,6)	0	0	0	5	-12	0	0	2	99	1
21	(0,2,5,3)	3	69	(0,3,4,6)	0	-3	0	0	-12	0	5	10	99	1
22	(0,2,5,4)	4	63	(0,1,3,4,6)	0	-3	0	5	-7	0	5	0	90	1
23	(0,2,5,6)	6	63	(0,1,3,4)	0	-3	0	5	-12	0	0	10	91	1
24	(0,2,6,1)	1	61	$(\tilde{I}/(1,6))$	0	0	0	5	-12	3	0	10	89	2
25	(0,2,6,3)	3	71	$(\tilde{I}/(3,6))$	0	-3	0	1	-12	3	0	11	103	2
26	(0,2,6,4)	4	58	$(\tilde{I}/(4,6))$	0	-3	0	5	-12	3	0	11	85	2
27	(0,2,6,5)	5	54	$(\tilde{I}/(5,6))$	0	-3	0	5	-12	0	0	10	79	1
28	(0,4,5,1)	1	56	(0,2,3,4,6)	0	0	-6	5	-4	0	5	0	70	2
29	(0,4,5,2)	2	70	(0,1,3,4,6)	0	-3	0	5	-4	0	5	9	91	2
30	(0,4,5,3)	3	59	(0,1,2,4,6)	0	-3	-6	0	-4	0	5	8	88	2
31	(0,4,5,4)	4	53	0,1,2,3,4,6	0	-3	-6	5	-1	0	5	0	81	2
32	(0,4,5,6)	6	53	0,1,2,3,4,6	0	-3	-6	5	-4	0	0	8	80	2
33	(0,4,6,1)	1	61	0,2,3,4,5	0	0	-6	5	-4	3	0	8	88	2
34	(0,4,6,2)	2	55	0,1,2,3,4,5	0	-3	-1	5	-4	3	0	0	87	2
35	(0,4,6,3)	3	74	0,1,2,4,5	0	-3	-6	0	-4	3	0	10	102	2
36	(0,4,6,4)	4	61	(0,1,2,3,4,5)	0	-3	-6	5	0	3	0	1	84	2
37	(0,4,6,5)	5	57	0,1,2,3,4	0	-3	-6	5	-4	0	0	8	78	2
38	(0,6,2,1)	1	58	0,3,4,5,6	0	0	0	5	-12	3	2	8	88	2
39	(0,6,2,3)	3	60	0,1,4,5,6	0	-3	0	0	-12	3	2	10	86	2
40	(0,6,2,4)	4	46	0,1,4,5,6	0	-3	0	5	-7	3	2	0	66	2
41	(0,6,2,5)	5	53	0,1,3,4,6	0	-3	0	5	-12	0	2	8	76	1
42	(0,6,2,6)	6	44	0,1,3,4,5	0	-3	0	5	-12	3	0	7	64	2
43	(0,25,1,3)	3	78	(0,4,6)	0	0	0	0	-12	0	5	7	113	1
44	(0,2,5,1,4)	4	84	(0,3,4,6)	0	0	0	5	-12	0	5	0	122	1
45	(0,2,5,3,1)	1	79	(0,4,6)	0	0	0	5	-12	0	5	7	113	1
46	(0,2,5,3,4)	4	89	(0,1,4,6)	0	-3	0	0	-2	0	3	0	126	2
47	(0,2,5,3,6)	6	104	(0,1,4,6)	0	-3	0	0	-12	0	4	11	147	2
48	(0,2,5,4,3)	3	83	(0,1,4,6)	0	-3	0	0	-7	0	5	5	117	2
49	(0,2,5,4,6)	6	85	(0,1,3,4)	0	-3	0	5	-7	0	0	5	120	2
50	(0,2,5,6,1)	1	88	(0,3,4)	0	0	0	5	-12	0	0	7	125	1
51	(0,2,5,6,3)	3	98	(0,1,3,4)	0	-3	0	4	-12	0	0	11	139	2
52	(0,2,5,6,4)	4	83	(0,1,3,4)	0	-3	0	5	-2	0	0	0	118	2
53	(0,2,6,5,1)	1	77	(0,3,4)	0	0	0	5	-12	0	0	7	111	1
54	(0,2,6,5,3)	3	78	(0,1,3,4)	0	-3	0	4	-12	0	0	11	111	2
55	(0,2,6,5,4)	4	72	(0,1,3,4)	0	-3	0	5	-2	0	0	0	104	2
56	(0,4,5,6,1)	1	78	(0,2,3,4)	0	0	-6	5	-4	0	0	5	114	2
57	(0,4,5,6,2)	2	68	(0,1,3,4)	0	-3	0	5	-4	0	0	2	103	2
58	(0,4,5,6,3)	3	88	(0,1,2,3,4)	0	-3	-6	2	-4	0	0	11	128	2

№	\tilde{U}^s	l^s	$\theta(\tilde{U}^s)$	$\hat{j}^{s,-}$	Объемы грузов в пунктах $g_k(\tilde{U}^s)$						G^s	$f(\tilde{U}^s)$	$\psi(\tilde{U}^s)$	
					0	1	2	3	4	5				6
59	(0,4,5,6,4)	4	75	(0,1,2,3)	0	-3	-6	5	0	0	0	4	110	2
60	(0,25,1,3,4)	4	98	(0,4,6)	0	0	0	0	-5	0	5	0	140	1
61	(0,25,1,3,6)	6	113	(0,4,6)	0	0	0	0	-12	0	1	11	161	3
62	(0,25,1,4,3)	3	104	(0,4,6)	0	0	0	0	-10	0	5	5	149	3
63	(0,25,1,4,6)	6	106	(0,3,4)	0	0	0	5	-10	0	0	5	143	1
64	(0,25,1,6,3)	3	127	(0,3,4)	0	0	0	0	-10	0	5	11	181	3
65	(0,25,1,6,4)	4	115	(0,4,6)	0	0	0	5	-5	0	0	0	173	3
66	(0,25,3,1,4)	4	95	(0,4,6)	0	0	0	0	-5	0	5	0	136	1
67	(0,25,3,1,6)	6	104	(0,4,6)	0	0	0	0	-12	0	1	11	147	1
68	(0,2,5,6,1,3)	3	98	(0,3,4)	0	0	0	1	-12	0	0	11	139	1
69	(0,2,5,6,1,4)	4	104	(0,3,4)	0	0	0	5	-5	0	0	0	150	1
70	(0,2,5,6,1,3,6)	6	120	(0,4)	0	0	0	0	-5	0	0	5	170	3
71	(0,2,5,1,4,6,3)	3	128	(0,4)	0	0	0	0	-10	0	0	10	173	3
72	(0,2,5,3,1,4,6)	6	117	(0,4)	0	0	0	0	-5	0	0	5	166	1
73	(0,2,5,3,1,6,4)	4	126	(0,3,4)	0	0	0	0	-1	0	1	0	177	1
74	(0,2,5,6,1,3,4)	4	118	(0,3,4)	0	0	0	1	-1	0	0	0	166	1
75	(0,25,6,1,4,3)	3	124	(0,4)	0	0	0	0	-5	0	0	5	177	1
76	(0,2,5,1,4,6,3,4)	4	148	(0)	0	0	0	0	0	0	0	0	200	3
77	(0,2,5,3,1,4,6,4)	4	139	(0)	0	0	0	0	0	0	0	0	196	1
78	(0,2,5,3,1,6,4,6)	6	148	(0,4)	0	0	0	0	-1	0	0	1	207	3
79	(0,25,6,1,3,4,3)	3	138	(0,4)	0	0	0	0	-1	0	0	1	193	1
80	(0,25,6,1,4,3,4)	4	144	(0,4)	0	0	0	0	0	0	0	0	204	3
81	(0,2,5,3,1,4,6,4,0)	0	156	(0)	0	0	0	0	0	0	0	0	220	1
81	(0,2,5,3,1,6,4,6,4)	4	170	(0)	0	0	0	0	0	0	0	0	247	3
82	(0,25,6,1,3,4,3,4)	4	156	(0)	0	0	0	0	0	0	0	0	220	3

Yu. A. Zack, Dokt.-Ing. Deutschlang, Aahen, Deutschland, e-mail: yuri_zack@hotmail.com

A Problem of Constructing Admissible and Optimal Routes of Delivery

The problem of creation of optimum routes of the movement of cargo vehicles, important for practical applications, in the conditions of restrictions on terms of service of points and on their loading capacity is considered. As criteria of an optimality total costs of implementation of the schedule of cargo delivery are considered. On the basis of the established properties of admissible and optimum solutions of a task rules of elimination of inadmissible and nonoptimal routes are formulated. Algorithms of the solution of a task are developed by the modified methods of dynamic programming and branches and borders which are illustrated with a numerical example. Results of computing experiments are given.

Keywords: cargo vehicles, optimum routes of the movement, restrictions on the terms of service, loading capacity and volumes of freights, consecutive algorithms of optimization, sequential optimization algorithms, the modified branch and bound method

References

1. Dantzig G. B., Ramser R. H. The Truck Dispatching Problem. *Management Science*. 1959. N. 6. P. 80–91.
2. Zack Yu. A. Obochennaja zadacha kommivojazhera i jeje prikladnije aspekti. *Avtomatika i telemekhanika*. 1976. N. 10. P. 120–132.
3. Zack Yu. A. Prikladnije zadachi teorii raspisanij i marchrutizazii perevosok. M.: URSS, Librokom, 2012. 393 p.
4. Bronshtein E. M., Zaiko T. A. Determinirovannije optimisazionnije zadachi transportnoi logistiki. *Avtomatika i telemekhanika*. 2010. N. 10. P. 133–147.
5. Vehicle Routing Problem. URL: <http://neo.lcc.uma.es/vrp/vehicle-routing-problem>. Solution Methods for VRP. URL: <http://neo.lcc.uma.es/vrp/solution-methods/>.
6. Bronshtein E. M., Davletov A. Nestazionarnaja zadacha upravlenija marchrutom transportnogo sredstva. *Problemi upravlenija*. 2014. N. 3. P. 23–28.
7. Malandraky C., Daskin M. S. Time Dependent Vehicle Routing Problems Formulations, Properties and Heuristic Algorithms. *Transportation Science*. 1992. Vol. 26. P. 185–200.
8. Crevier B., Cordeau J.-F., Laporte G. The multi-depot vehicle routing problem with inter-depot routes. *European Journal of Operational Research*. 2007. Vol. 176. P. 756–773.
9. Stegers E. A Solution Method for Vehicle Routing problems with Time-Dependent Travel Times. Delft: Delft University of Technology, 2009. 83 p.
10. Ralphs T. K., Kopman L., Pulleyblank W. R., Trotter L. E. On the Capacitated Vehicle Routing Problem. *Math. Program. Ser. B*. 2003. Vol. 94. P. 343–359.

П. С. Панкратьев, аспирант, e-mail: scud33@inbox.ru,
В. А. Шакиров, канд. техн. наук, доц., e-mail: mynovember@mail.ru,
Братский государственный университет, Россия, Братск

Двухуровневый многокритериальный анализ пунктов строительства гидроэлектростанции

Формализуется проблема принятия решений по выбору пунктов строительства гидроэлектростанции. Проблема отличается необходимостью анализа двух уровней альтернатив — пунктов строительства и вариантов станций, дефицитом информации по альтернативам и многочисленными критериями из оценки. Формулируются требования к методам анализа альтернатив на двух уровнях. Для решения задачи предлагается использовать метод анализа иерархий и метод многокритериальной теории полезности. Предлагается методика решения задачи в такой постановке. Проводится выбор пункта строительства гидроэлектростанции на реке Индигирке.

Ключевые слова: многокритериальный анализ, теория полезности, метод анализа иерархий, размещение гидроэлектростанции

Введение

Процесс принятия решений по размещению гидроэлектростанций (ГЭС) имеет ряд особенностей. Во-первых, решения по указанной проблеме имеют долговременные социальные, экологические, экономические, биологические последствия, что требует тщательного многокритериального анализа. Во-вторых, анализ проводится в условиях неопределенности исходной информации и будущих условий, так как период времени от начальных этапов проектирования до ввода в эксплуатацию может достигать 15 лет. В-третьих, проблемы размещения ГЭС относятся к слабоструктурированным.

По ряду критериев получение точной количественной оценки не всегда возможно в силу ограниченности финансовых и временных ресурсов, поэтому используют качественные оценки, полученные от экспертов. В силу отмеченных особенностей, других сложностей, присущих системным проблемам, задачу разбивают на этапы. Например, на первом этапе намечаются перспективные пункты строительства ГЭС, на втором этапе уточняется ее мощность, площадка строительства, технические параметры. Этапы рассматриваются отдельно, в результате выбранный на первом этапе вариант решения может иметь набор неэффективных альтернатив на втором этапе. Повторный анализ ведет к сильной нагрузке лица, принимающего решения (ЛПР), потере финансовых и временных ресурсов [1–3], поэтому предлагается уже при выборе пунктов строительства ГЭС учитывать оценки возможных в этих пунктах дальнейших решений.

Постановка задачи

При сравнении пунктов строительства предлагается выделить два уровня

альтернатив. Альтернативами первого уровня (АПУ) являются пункты строительства, альтернативами второго уровня (АВУ) — возможные в пункте варианты ГЭС. Например, одной из важнейших характеристик ГЭС является нормальный подпорный уровень (НПУ) водохранилища ГЭС, который определяет мощность станции, капитальные вложения, площадь затопления земель и т. д. Каждый пункт в зависимости от гидрологических характеристик соответствующего района обладает определенным диапазоном возможных НПУ. Именно поэтому при сравнении пунктов целесообразно сразу включить в анализ оценки возможных к реализации вариантов электростанции.

Сформулируем задачу следующим образом. Пусть $A = \{a_1, a_2, \dots, a_k\}$ — множество АПУ, которые оцениваются по множеству критериев $F = F_1 \cup G$. Подмножество критериев $F_1 = \{f_1, f_2, \dots, f_s\}$ служит для оценки альтернатив только первого уровня, подмножество критериев $G = \{g_1, g_2, \dots, g_p\}$ — для оценки альтернатив и первого, и второго уровней. Каждому элементу a_i множества A ставится в соответствие множество возможных АВУ $B_i = \{b_{i1}, b_{i2}, \dots, b_{im}\}$. Необходимо упорядочить по предпочтению альтернативы множества A с учетом многокритериальных оценок альтернатив множеств B_i (рис. 1).

Особенностью задачи является то, что выбор лучших АВУ, лучшей АПУ проводится в различных

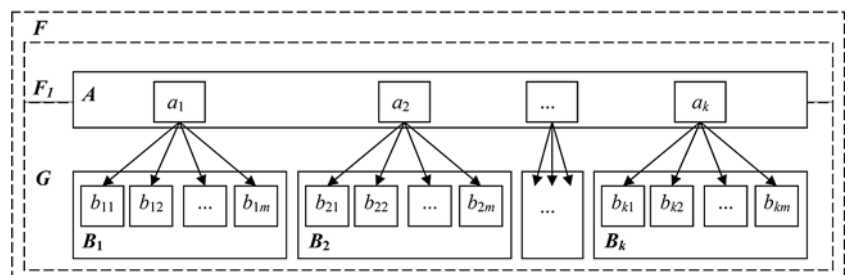


Рис. 1. Двухуровневая структура проблемы выбора пункта строительства ГЭС

условиях — различное число альтернатив, описание критериев, поэтому целесообразно использовать два метода многокритериального анализа.

АВУ, как правило, характеризуется критериями, по которым может быть дана количественная оценка — мощность ГЭС, стоимость ГЭС, площадь затопленных земель. Число альтернатив второго уровня потенциально не ограничено, так как шаг изменения оценки по критериям внутри диапазона может быть выбран очень малым. Соответственно необходим метод, позволяющий проводить многокритериальную оценку в условиях большого числа альтернатив по критериям с количественным описанием. Этим требованиям отвечает метод многокритериальной теории полезности (Multiattribute utility theory (MAUT)). Данный метод широко применялся на практике [4, 5].

АПУ также могут характеризоваться количественными критериями (например, сейсмичность в районе строительства), но в большей степени для описания альтернатив привлекаются критерии, по которым может быть дано только качественное описание, например, условия строительства, перспективы развития района. А по некоторым критериям могут быть даны только сравнительные оценки для альтернатив, например, воздействие на ихтиофауну. Число альтернатив первого уровня, как правило, не превышает 10. Для описанных условий эффективен метод анализа иерархий (МАИ). МАИ также применялся при решении важных практических задач [8].

Таким образом, предлагается следующий подход к выбору пункта для строительства ГЭС (рис. 2). На первом этапе с помощью картографических материалов намечаются пункты (АПУ), в каждом из которых проводятся водноэнергетические расчеты и определяются диапазоны возможных НПУ (АВУ). На втором этапе формируется иерархия целей и критериев с разделением их на две группы — F_1 и G . На третьем этапе для каждого пункта по критериям множества G методом MAUT выбирается лучшая АВУ. На четвертом этапе пункты сравниваются методом анализа иерархий в предположении, что именно лучшая АВУ будет реализована в пункте. При сравнении пунктов используются критерии F_1 и G .

Предлагаемый подход позволяет при сравнении пунктов учесть качество возможных последующих решений, т. е. учесть оценку вариантов станций, которые будут в дальнейшем реализованы.



Рис. 2. Двухуровневый подход к выбору пунктов строительства ГЭС

Пример использования двухуровневого подхода

Рассмотрим предлагаемый подход к многокритериальному выбору наиболее перспективного пункта строительства ГЭС на примере реки Индигирки, протекающей на севере республики Саха (Якутия). Ситуационный план района представлен на рис. 3.

В исследовании будет рассмотрен участок реки, расположенный в перспективных для развития Момском и Оймяконском улусах. Достаточно благоприятные гидрологические условия района, сосредоточение месторождений золота, сурьмы, вольфрама, каменного угля позволяют провести анализ нескольких вариантов строительства ГЭС. На реке Индигирке для анализа было выбрано пять пунктов возможного строительства ГЭС, два из которых — уже были намечены в СССР Институтом физических проблем Севера (ИФТПС СО РАН) [6].

Пункт П1 намечен вблизи поселка Хонуу, расположенного в устье реки Момы. Перспективу поселка определяет стратегия развития железнодорожного транспорта в Российской Федерации до 2030 г., утвержденная распоряжением правительства РФ от 17 июня 2008 г. № 877-р, в соответствии с которой запланировано строительство стратегической железнодорожной линии Якутск—Мома—Магадан. Кроме того, от поселка Хонуу Индигирка судоходна. Таким образом, в перспективе поселок может стать крупным транспортным узлом. Рост



Рис. 3. Ситуационный план района

Таблица 1

Результаты водноэнергетических расчетов

Пункт (АПУ)	АВУ	НПУ, м	Среднесуточная гарантированная мощность, МВт	Площадь затопления, км ²
П1 (АПУ 1)	АВУ 11	30	3,601	21,320
	АВУ 12	40	11,771	30,920
	АВУ 13	50	24,843	44,390
	АВУ 14	60	40,265	61,730
П2 (АПУ 2)	АВУ 21	30	1,594	6,760
	АВУ 22	40	10,912	41,650
	АВУ 23	50	28,803	73,000
П3 (АПУ 3)	АВУ 31	30	8,887	50,000
	АВУ 32	40	24,089	79,400
П4 (АПУ 4)	АВУ 41	20	4,594	51,120
	АВУ 42	30	17,644	82,760
	АВУ 43	40	38,136	115,110
	АВУ 44	50	74,860	201,210
	АВУ 45	60	129,735	269,860
П5 (АПУ 5)	АВУ 51	20	4,712	50,800
	АВУ 52	30	16,241	109,450
	АВУ 53	40	41,935	193,550

электрических нагрузок будет обусловлен строительством и обслуживанием железной дороги.

Пункт П2 был намечен ИФТПС СО РАН в горной речной долине у устья реки Луч. Территория является равноудаленной от поселков Предпорожный и Хонуу.

Пункт П3 намечен вблизи поселка городского типа Предпорожный, упраздненного в 2007 г. Вблизи расположены месторождения золота, угля. Расположение между поселками Хонуу и Усть-Нера позволяет влиять на развитие двух районов.

Пункт П4 намечен вблизи поселка городского типа Усть-Нера, расположенного вблизи богатейших месторождений золота, сурьмы, вольфрама. Через поселок проходит федеральная автомобильная дорога Р504 "Колыма", связывающая Якутск и Магадан.

Пункт П5 — второй, намеченный ИФТПС СО РАН, — расположен на территории, богатой месторождениями рудного и россыпного золота.

На первом этапе (см. рис. 2) после формирования АПУ были проведены водноэнергетические расчеты, целью которых является определение возможных площадей затопления, гарантированной выработки электроэнергии и среднесуточной мощности ГЭС [1]. Результаты расчетов представлены в табл. 1. Последовательность расчетов и промежуточные результаты для трех пунктов приведены в работе [1].

На втором этапе формируется иерархия целей и критериев (рис. 4). На рисунке показано разделение критериев на две группы: К1, К2, К3 — критерии множества G (служат для оценки вариантов ГЭС и пунктов), критерии К4...К7 — критерии множества F_1 (служат только для оценки пунктов).

На третьем этапе исследования проводится выбор лучших АВУ (вариантов ГЭС) для возможных пяти пунктов (П1...П5). Для этого используется метод МАУТ, который имеет аксиоматическое обоснование. При выполнении условий-аксиом математически доказывается, что существует скалярная функция полезности $u(y)$, которая ставит в соответствие каждой альтернативе число, отражающее ее полезность [4]. Выполнение аксиом отражает готовность ЛПР сравнивать, ранжировать альтернативы, ухудшать оценки по одному критерию для улучшения оценки по другому.

Если рассматривается выбор в условиях определенности, функцию полезности $u(y)$ называют функцией ценности $v(y)$ [4, 9]. Так как отсутствует объективная или субъективная информация о распределениях вероятности изменения оценок по критериям, будем рассматривать решение задачи в условиях определенности и использовать функцию ценности $v(y)$. Выполнение условия взаимной независимости критериев по предпочтению [4, 9] позволяет получить многокритериальную функцию ценности (МФЦ) в аддитивном виде:

$$v(y) = v(y_1, y_2, \dots, y_n) = \sum_{i=1}^n k_i v_i(y_i),$$

где k_i — шкалирующий коэффициент критерия i ;

$\sum_{i=1}^n k_i = 1$; $v(y_i)$ — однокритериальная функция

ценности (ОФЦ); y_i — оценка АВУ по критерию i .

Для построения однокритериальных функций ценности проводится опрос ЛПР, в ходе которого



Рис. 4. Иерархия целей и критериев проблемы выбора пункта строительства ГЭС

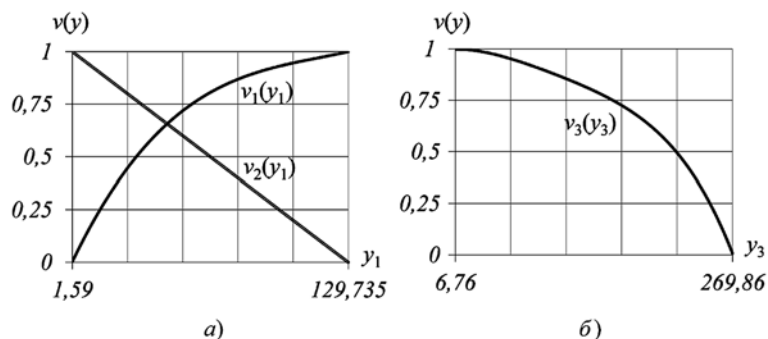


Рис. 5. ОФЦ критериев:
 а — $v_1(y_1)$ — среднесуточная мощность ГЭС, МВт; $v_2(y_1)$ — стоимость строительства ГЭС, о.е., б — $v_3(y_3)$ — площадь затопляемых земель, км²

определяется ценность $v(y_i)$ для всего возможного диапазона критериальных оценок y_i . Построенные ОФЦ представлены на рис. 5. Стоимость строительства ГЭС определялась из соотношения 200 тыс. руб. за 1 кВт, поэтому ОФЦ, устанавливающая ценность для вариантов мощности, $v_1(y_1)$ и ОФЦ, устанавливающая ценность для стоимости ГЭС, $v_2(y_1)$ построены от одной переменной y_1 — среднесуточной мощности ГЭС.

Далее, в ходе диалоговых процедур [4, 5] устанавливаются отношения важности между критериями и решается система уравнений. В результате определяются шкалирующие коэффициенты k и многокритериальная функция ценности:

$$v(y_1, y_2, y_3) = k_1 v_1(y_1) + k_2 v_2(y_1) + k_3 v_3(y_3).$$

В исследовании получена МФЦ:

$$v(y_1, y_2, y_3) = 0,403v_1(y_1) + 0,481v_2(y_1) + 0,116v_3(y_3).$$

С помощью МФЦ могут быть ранжированы все варианты ГЭС для каждого пункта. На рис. 6 представлены графики зависимости площади затопления земель от среднесуточной мощности анализируемых пяти пунктов. На рисунке также отображены линии фиксированной ценности, вычисленной с помощью полученной МФЦ для всех значений ОФЦ. Из рисунка видно, какие АБУ достигают наибольшей ценности для каждого пункта; они и будут отображены для третьего этапа сравнения АПУ: для пункта 1 — вариант ГЭС с НПУ 60 м, для пункта 2 — 50 м; пунктов 3, 4, 5 — 40 м.

На четвертом этапе применяется метод анализа иерархий. Метод использует относительную шкалу сравнения (табл. 2) [7].

В соответствии с методом анализа иерархий ЛПР проводит попарные сравнения альтернатив по критериям с помощью шкалы относительной важности, формируя матрицы парных сравнений. Для этого попарно сравниваются альтернативы из строки с каждой альтернативой из столбца (табл. 3). Оценки шкалы относительной важности c_{ij} вписываются в ячейки, образованные пересечением соответствующей строки и столбца.

Так формируются матрицы парных сравнений альтернатив. Затем аналогично проводится парное сравнение критериев, формируется матрица парных сравнений критериев [7]. Вычислив собственные векторы матриц парных сравнений и проведя нормирование элементов собственных векторов, можно получить веса критериев и альтернатив. Определение собственного вектора матрицы $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ проводилось по известному выражению [7]

$$\lambda_i = \frac{1}{n} \sqrt[n]{\prod_{j=1}^n a_{ij}},$$

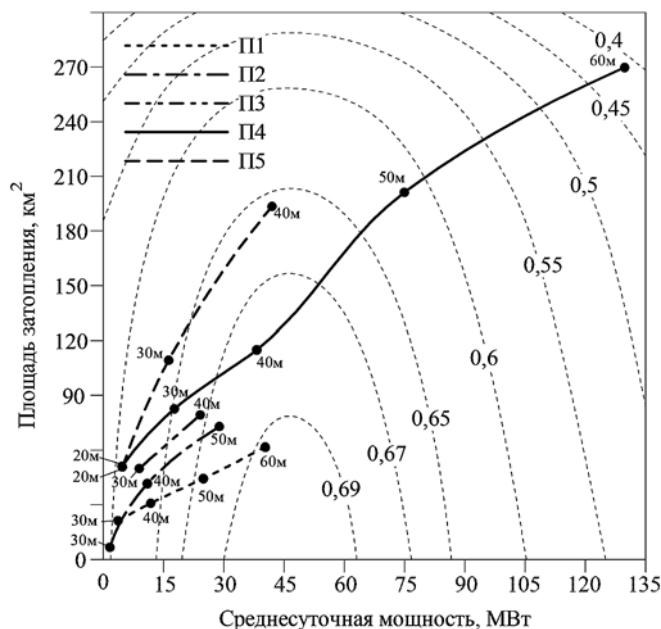


Рис. 6. Зависимости площади затопления от среднесуточных мощностей для каждого пункта и линии фиксированных многокритериальных значений ценностей

Таблица 2

Шкала относительной важности

Уровень важности	Количественное значение
Равная важность	1
Умеренное превосходство	3
Значительное превосходство	5
Явное превосходство	7
Абсолютное превосходство	9

Таблица 3

Заполнение матрицы парных сравнений

Критерии	f_1	f_2	...	f_s
f_1	1	c_{12}	...	c_{1s}
f_2	c_{21}	1	...	c_{2s}
...	1	...
f_s	c_{s1}	c_{s2}	...	1

Таблица 4

Матрица парных сравнений по критерию "Среднесуточная мощность ГЭС"

Пункт	П1	П2	П3	П4	П5	Вес
П1	1	7	8	3	0,500	0,310
П2	0,143	1	2	0,200	0,125	0,048
П3	0,125	0,500	1	0,167	0,111	0,033
П4	0,333	5	6	1	0,250	0,154
П5	2	8	9	4	1	0,456

Таблица 5

Матрица парных сравнений критериев

Критерий	K1	K2	K3	K4	K5	K6	K7	Вес
K1	1	2	3	0,500	5	0,500	5	0,179
K2	0,500	1	2	0,333	4	0,333	4	0,116
K3	0,333	0,500	1	0,250	3	0,250	3	0,076
K4	2	3	4	1	6	1	6	0,280
K5	0,200	0,250	0,333	0,167	1	0,167	1	0,036
K6	2	3	4	1	6	1	6	0,280
K7	0,200	0,25	0,333	0,167	1	0,167	1	0,036

Таблица 6

Оценки альтернатив

Альтернатива	Оценка
АПУ1 (П1)	0,293
АПУ2 (П2)	0,104
АПУ3 (П3)	0,145
АПУ4 (П4)	0,209
АПУ5 (П5)	0,250

где λ_i — элемент собственного вектора матрицы, соответствующий альтернативе или критерию i ; a_{ij} — оценка шкалы парных сравнений альтернатив или критериев i и j ; n — число альтернатив или критериев.

Вес критерия w_i или альтернативы v_{ij} определяется путем нормирования элементов собственного вектора. Например, вес критерия определяется формулой

$$w_i = \frac{\lambda_i}{\sum_{i=1}^n \lambda_i} \quad (1)$$

Оценки альтернатив с учетом всех критериев получают по выражению

$$V_j = \sum_{i=1}^n w_i v_{ij} \quad (2)$$

где V_j — показатель качества j -й альтернативы; w_i — вес i -го критерия; v_{ij} — вес j -й альтернативы по i -му критерию.

При сравнении АПУ по критериям "Среднесуточная мощность ГЭС", "Стоимость строительства", "Площадь затопления" участвуют выбранные ранее

лучшие АБУ. Пример матрицы парных сравнений АПУ по критерию "Среднесуточная мощность ГЭС" и полученные по формуле (1) веса представлены в табл. 4. Аналогично проводилось сравнение АПУ по критериям K4—K7.

Матрица парных сравнений критериев, а также полученные по выражению (1) веса представлены в табл. 5. Многокритериальные оценки альтернатив, полученные по выражению (2), представлены в табл. 6.

Анализ полученных результатов

Итак, с позиций разработанной иерархии целей и критериев, назначенных оценок при парных сравнениях лучшим является пункт П1.

Строительство ГЭС в пункте П1 позволит обеспечить развитие транспортного узла Хонуу, решить вопросы электроснабжения объектов железной дороги при выполнении стратегии развития железнодорожного транспорта до 2030 г. Пункт обеспечивает наименьшую площадь затопления, в то же время сравнительно высокую по сравнению с конкурентными вариантами мощность (вторую по величине), расположен в зоне с низкой относительно других створов сейсмичностью. Но по такому критерию как "Воздействие на ихтиофауну" пункт имеет худшую оценку. Низкая оценка альтернативы и по стоимости строительства. Тем не менее, с учетом выраженных предпочтений, пункт П1 имеет самую высокую многокритериальную оценку.

Также высокую оценку получил пункт П5, имеющий высокие перспективы для развития района, богатого месторождениями цветных и благородных металлов. Пункт имеет наименьшее воздействие на ихтиофауну, лучший вариант ГЭС по мощности. Однако пункт имеет низкую сравнительную оценку по условиям строительства, стоимости ГЭС, самую низкую оценку по площади затопления среди рассматриваемых на втором уровне альтернатив.

Заключение

Новизна предлагаемого подхода заключается в многокритериальном рассмотрении проблемы на двух уровнях с соответствующим формированием двух уровней альтернатив. Каждая альтернатива первого уровня может иметь возможные реализации в виде альтернатив второго уровня. Альтернативы первого и второго уровней имеют различную степень детализации описаний последствий по критериям, что требует применения различных подходов к многокритериальной оценке. Предложена методика для решения задачи в такой постановке.

Предложенный многокритериальный подход к выбору пунктов ГЭС может проводиться на ранних этапах инженерных изысканий, в предпроектных исследованиях, когда известны лишь ориентировочные данные в отношении сооружаемой ГЭС,

в целях определения наиболее предпочтительных мест для дальнейшего, более детального анализа. Особенностью подхода является возможность избежать точной оценки альтернатив по некоторым критериям, ограничившись лишь выражением превосходства альтернатив по отношению друг к другу.

Список литературы

1. Панкратьев П. С., Шакиров В. А. Многокритериальный выбор створа гидроэлектростанции на реке Индигирке в республике Саха (Якутия) // Системы. Методы. Технологии. 2012. № 3 (15). С. 71–80.
2. Шакиров В. А., Панкратьев П. С. Многокритериальный двухуровневый подход к выбору лучшей альтернативы в рамках слабоструктурированной проблемы // Вестник Воронежского

государственного университета. Сер. "Системный анализ и информационные технологии". 2013. № 2. С. 118–127.

3. Шакиров В. А., Панкратьев П. С. Поддержка принятия решений на стадии предпроектных исследований на основе двухуровневого многокритериального анализа // Прикладная информатика. 2013. № 6(48). С. 111–121.
4. Кини Р. Л., Райф Х. Принятие решений при многих критериях: предпочтения и замещения: Пер. с англ. / Под ред. И. Ф. Шахнова. М.: Радио и связь, 1981. 580 с.
5. Кини Р. Л. Размещение энергетических объектов: выбор решений: Пер. с англ. М.: Энергоматиздат, 1983. 320 с.
6. Ноговицын Д. Д. Водные ресурсы Якутской АССР и их использование. Якутск: Изд. ЯФ СО АН СССР, 1985. 124 с.
7. Саати Т. Л. Принятие решений. Метод анализа иерархий: Пер. с англ. М.: Радио и связь, 1993. 316 с.
8. СП 14.13330.2011. Строительство в сейсмических районах.
9. Черноруцкий И. Г. Методы принятия решений. СПб.: БХВ-Петербург, 2005. 416 с.

P. S. Pankratiev, Post Graduate Student, e-mail: scud33@inbox.ru,
V. A. Shakirov, Assistant Professor; e-mail: mynovember@mail.ru,
Bratsk State University, Russia, Bratsk

Two-Level Multiattribute Analysis of Hydroelectric Station Sites

In this paper, decision-making problem of choice of hydroelectric station sites is formalized. The problem characterized by the need to analyze two alternative levels, which are sites and variants of stations, by shortage of information on alternatives and multiple criteria. Requirements for methods of alternative analysis for two levels are formulated. To solve the problem analytic hierarchy process and multi-attribute utility theory are provided. A method of solving the problem in such a setting is provided. Siting of hydroelectric power plant on the Indigirka River is performed.

Keywords: multiattribute analysis, utility theory, analytic hierarchy process, siting of hydroelectric station

References

1. Pankrat'ev P. S., Shakirov V. A. Mnogokriterial'nyi vybor stvora gidroelektrostantsii na reke Indigirke v respublike Saha (Jakutija). *Sistemy. Metody. Tehnologii*. 2012. N. 3(15). P. 71–80.
2. Shakirov V. A., Pankrat'ev P. S. Mnogokriterial'nyi dvuhurovnevyy podhod k vyboru luchshej al'ternativy v ramkah slabostrukturovannoj problemy. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Ser. "Sistemnyj analiz i informacionnye tehnologii"*. 2013. N. 2. P. 118–127.
3. Shakirov V. A., Pankrat'ev P. S. Podderzhka prinjatja reshenij na stadii predproektnyh issledovanij na osnove dvuhurovnevnogo mnogokriterial'nogo analiza. *Prikladnaja informatika*. 2013. N. 6(48) 2013. P. 111–121.

4. Kini R. L., Raifa H. *Prinjatje reshenij pri mnogih kriteriiah: predpochtenija i zameshhenija*: Per. s angl. Pod red. I. F. Shahnova. M.: Radio i svjaz', 1981. P. 580.
5. Kini R. L. *Razmeshhenie jenergeticheskij ob'ektov: vybor reshenij*: Per. s angl. M.: Jenergomatizdat, 1983. P. 320.
6. Nogovicyn D. D. *Vodnye resursy Jakutskoj ASSR i ih ispol'zovanie*. Jakutsk: izd JaF SO AN SSSR, 1985. P. 124.
7. Saati T. L. *Prinjatje reshenij. Metod analiza ierarhij*: Per. s angl. M.: Radio i svjaz', 1993. P. 316.
8. СП 14.13330.2011. *Stroitel'stvo v sejsmicheskijh rajonah*.
9. Chernoruckij I. G. *Metody prinjatija reshenij*. SPb.: BHV-Peterburg, 2005. 416 p.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОРГАНИЗАЦИОННЫХ И СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ СИСТЕМАХ INFORMATION TECHNOLOGIES IN THE ORGANIZATIONAL AND SOCIO-ECONOMIC SYSTEMS

УДК 001.891.57

О. Н. Андрейчикова, д-р техн. наук, вед. науч. сотр.

Центральный экономико-математический институт Российской академии наук (ЦЭМИ РАН),
e-mail: alexandrol@mail.ru

А. А. Вавилов, магистр,

Национальный исследовательский университет "Московский физико-технический институт" (МФТИ),

А. Н. Козырев, д-р экон. наук, зав. лаб.,

Центральный экономико-математический институт Российской академии наук (ЦЭМИ РАН)

Исследование поведения потребителей на рынке программного обеспечения*

Приведены результаты исследования поведения индивидуальных потребителей программного обеспечения с применением инструмента имитационного моделирования AnyLogic. Выявлены главные факторы, определяющие поведение пользователей, и ключевые точки принятия решений. Построена агентная модель, на которой проведена серия вычислительных экспериментов. Полученные результаты демонстрируют хорошее соответствие наблюдаемым данным и могут быть использованы для анализа возможных стратегий распространения программных продуктов.

Ключевые слова: имитационное моделирование, поведение потребителей, компьютерное пиратство, агентная модель, рынок программных продуктов

Введение

Рынок программного обеспечения (ПО) быстро растет во всем мире. По данным отчетов BSA (*Business Software Alliance*) и IDC (*International Data Corporation*) с 2008 по 2012 гг. ежегодный прирост рынка информационных услуг в России превышал 25 % и был самым высоким в Европе [1–4]. Уровень пиратства в данной сфере постепенно снижается, но остается весьма высоким. Оценки уровня пиратства в сфере ПО, приведенные в отчете BSA за 2011 г. [3], показаны на рис. 1. По данным отчета BSA за 2013 г. [5] уровень пиратства в сфере ПО вырос на 1 % и достиг 43 %, в то время как для России он снизился с 63 до 62 %.

Причины компьютерного пиратства весьма разнообразны. Помимо высоких цен на лицензионные программы к ним можно отнести широкое распространение Интернета и локальных сетей, сложность и низкую эффективность правоприменительной

*Работа выполнена при поддержке гранта РНФ проект № 14-18-01999 "Измерения и прогнозы в цифровой экономике".

практики даже при строгом законодательстве в данной области, психологические аспекты потребления интеллектуальных продуктов. Основным инструментом, который применяется для выявления причин пиратства, являются социологические опросы. На рис. 2 (см. третью сторону обложки) приведена диаграмма, демонстрирующая результаты опроса,

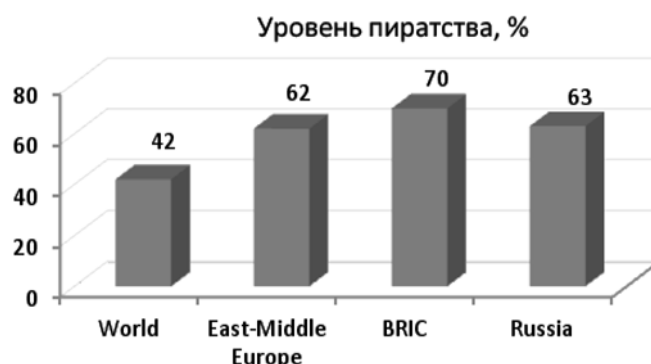


Рис. 1. Оценки уровня пиратства в сфере ПО (по данным отчета BSA 2011)

проведенного BSA в 2011 г., которая однозначно свидетельствует о более высокой склонности к пиратству российских потребителей ПО по сравнению со средним уровнем по всему миру. При этом основные отличия наблюдаются для категорий потребителей, допускающих периодическое использование нелегальных программных продуктов. Доля убежденных пиратов в России и в мире примерно совпадает, а вот доля принципиальных противников компьютерного пиратства в России почти в полтора раза ниже, чем в среднем по миру. Для выявления причин таких различий в данной работе мы пытались смоделировать поведение пользователей ПО при поиске и приобретении необходимых программных продуктов.

По данным аналитиков, основным источником контрафактного ПО в настоящее время является сеть Интернет. При этом примерно 76 % всех сайтов и файлообменных сетей, с которых скачивается ПО, пытаются установить на пользовательский компьютер вредоносные программы [3, 6]. Вероятно, принимая решения об установке пиратских программ, потребители сознательно принимают высокий риск заражения своих компьютеров. Очевидно, что этот риск является платой за те выгоды, которые получает потребитель нелегального ПО. В табл. 1 систематизированы выгоды и издержки приобретателей нелегального ПО.

Заметим, что издержек и рисков здесь больше, чем выгод, следовательно, приоритеты выгод для пиратов существенно выше, чем приоритеты возможных негативных последствий. Получается, что возможность быстро найти и приобрести бесплат-

ную (или почти бесплатную) пиратскую программу перевешивает страх нарушения закона и заражения вирусами, а также заставляет пользователя мириться с низким качеством продукта и прочими неудобствами. Данные опросов российских пользователей подтверждают эти предположения, в частности, только 23 % респондентов отвергают пиратские программы по убеждениям, и всего 19 % отказываются от пиратства ввиду риска быть пойманными [4].

Исследование поведения потребителей ПО позволяет выявить механизмы принятия решений и предложить эффективные меры для снижения уровня пиратства в России. Этот уровень является своеобразным индикатором социально-экономических проблем государства. Высокие значения этого индикатора свидетельствуют о бедности населения и о высоком спросе на ПО. Следует заметить, что мы рассматриваем ПО, которое не относится к сфере развлечений, поэтому высокий спрос на него говорит о стремлении общества к знаниям, что весьма достойно поощрения. Эффективные способы борьбы с пиратством должны обеспечивать определенный баланс между интересами разработчиков ПО (правообладателей) и доступностью программных средств для всех членов общества. В нашей работе [7] по исследованию эффективности различных способов борьбы с пиратством в сети Интернет показано, что ужесточение наказаний — далеко не лучший способ, поскольку имеет низкую селективность, приводит к появлению новых форм пиратства и ограничивает доступ к знаниям для значительной части общества.

1. Структура агентной модели

Исследование поведения потребителей программного обеспечения проводилось путем построения имитационной модели и последующих экспериментов с ней. В качестве основного инструмента моделирования использовался пакет имитационного моделирования AnyLogic 6 Professional. Агентная модель позволяет получить представление о свойствах системы на основе поведения ее составных частей — агентов, которые действуют в соответствии с заранее заданным набором правил [8]. Агентами являются потенциальные пользователи ПО, нуждающиеся в определенных программных продуктах, которые они пытаются найти на твердых носителях или в сети Интернет. В качестве пользователей рассматриваются физические лица, использующие компьютер для личных целей в домашних условиях. Корпоративные пользователи исключены из рассмотрения, в их среде в настоящее время уровень пиратства ниже, что объясняется более высокими рисками. В результате поиска агенты принимают решения о приобретении легального или пиратского программного продукта или остаются ни с чем. При этом существуют четыре

Таблица 1

Выгоды и издержки пользователей нелегальных программных продуктов

Критерии	Выгоды	Издержки и риски
Стоимость		
Качество продукта		
Доступность (затраты времени на поиск и приобретение)		
Разнообразие		
Возможности поддержки и обновления		
Законность приобретения		
Риск заражения вирусами		
Риск утечки персональных данных и мошенничества		
Нежелательная реклама		



Рис. 3. Структурная схема модели

возможных источника получения выбранной программы:

- лицензионный контент в Интернете (Интернет легальный);
- нелегальный контент в Интернете (Интернет пиратский);
- лицензионный диск с программой, купленный в магазине (твердый легальный);
- нелегальная копия диска (твердый пиратский).

Эксперименты с моделью позволяют выяснить, как изменяется число пиратов, легальных и неудовлетворенных пользователей при изменении ее параметров. В данном случае эксперимент — это последовательность действий, выполняемых программой при заданных начальных условиях. Завершение эксперимента происходит, когда все пользователи, число которых заранее определено, приняли решения о приобретении выбранного ПО. Структурная схема модели представлена на рис. 3.

Построенная модель, как и любая другая, не описывает все аспекты поиска и приобретения ПО, в частности, за рамками исследования остались механизмы функционирования поисковых систем, особенности дистрибуции и маркетинга программных средств, специфика Интернет-ресурсов и собственно контента и т. д. Тем не менее, набор параметров модели оказался достаточным для моделирования основных тенденций потребительского поведения на рынке программных средств.

Параметры контента. В данном исследовании многообразие контента было представлено шестью классами программных продуктов массового назначения: 1) операционные системы (версии Windows); 2) пакеты программ для офиса; 3) мультимедийные средства (проигрыватели); 4) файловые менеджеры; 5) антивирусы; 6) программы для работы с дисками. При проведении экспериментов каждый класс содержал по четыре конкретных программы.

Каждый экземпляр контента характеризуется ценой, полезностью, параметром качества пиратской копии, принадлежностью к определенному классу, текущей популярностью. Полезность оценивается в баллах по шкале от 0 до 10 и характеризует функциональные возможности, универсальность, гибкость и другие потребительские качества про-

граммного продукта. Качество пиратской копии — экспертная оценка на интервале (0, 1), соответствующая конкретному экземпляру программы. Текущая популярность — динамический параметр, который в ходе эксперимента увеличивается на единицу при каждом скачивании или покупке программного продукта.

Для каждого класса контента пользователи имеют бюджетные ограничения — *ClassBudgetLimit* (усредненная сумма, которую пользователи готовы потратить на приобретение программы из конкретного класса). Значения бюджетных ограничений, приведенные в табл. 2, определялись на основе данных опросов пользователей. Там же приведены усредненные цены и средние оценки качества пиратских копий программных продуктов для каждого класса.

Представление агентов. В экспериментах с моделью генерируются последовательности агентов, для каждого из которых определяются следующие параметры:

ClassNumber — класс контента, необходимый пользователю — равновероятный выбор из множества классов;

ContentNumber — выбранный пользователем программный продукт (либо равновероятный из экземпляров класса, либо выбор с вероятностью, пропорциональной текущей популярности);

Necessity — необходимость выбранного контента для пользователя, — $\sim N(0,8; 0,2^2)$ — нормальное распределение с математическим ожиданием 0,8 и стандартным отклонением 0,2;

BudgetLimit — верхняя граница бюджетного ограничения $\sim N(ClassBudgetLimit, ClassBudgetLimit/3)$ — нормальное распределение с математическим ожиданием, равным среднему бюджетному ограничению для класса контента и стандартным отклоне-

Таблица 2

Усредненные параметры классов программных продуктов

Наименование класса	Среднее бюджетное ограничение, <i>Class Budget Limit</i> , руб.	Усредненная цена программных продуктов класса, руб.	Качество пиратских копий (экспертная оценка по шкале 0—1)
Операционные системы (версии Windows)	3000	5400	0,7
Пакеты программ для офиса	2000	7200	0,65
Мультимедийные средства	700	1300	0,7
Файловые менеджеры	400	1000	0,7
Антивирусы	1000	1400	0,3
Программы для работы с дисками	400	1700	0,6

нием, равным одной трети бюджетного ограничения для класса;

Quality_required — требуемое качество контента;
 $\sim N(0,6; 0,2^2)$;

AttemptMax — максимальное число попыток поиска контента — равномерное распределение от 2 до 5;

ContactRate — число друзей пользователя, которым он может сообщить о случаях наказания за пиратство — равномерное распределение от 3 до 10;

AdaptionFraction — вероятность, с которой информация о возможном наказании может повлиять на решение пользователя, $\sim N(0,8; 0,2^2)$;

MultipleBL — мультипликатор бюджетного ограничения, равномерное распределение от 1 до 2 с шагом 0,1.

Конфигурация модели. Было исследовано несколько вариантов агентной модели:

1) базовый вариант основан на предположении, что пользователи действуют независимо друг от друга;

2) вариант, где пользователям доступна информация о числе приобретенных экземпляров каждой программы, которое интерпретируется как текущая популярность;

3) вариант, когда пользователи обмениваются информацией о наказаниях за пиратство.

В модели приняты следующие предположения:

- Поведение пользователей описывается диаграммой состояний, переходы между которыми происходят согласно заданным условиям и никак иначе. Диаграмма состояний приведена рис. 4.
- Параметры модели остаются постоянными в течение одного эксперимента.
- Каждый пользователь в итоге оказывается либо легальным (*L*), либо пиратом (*P*), либо неудовлетворенным (*X*). Неудовлетворенным считается пользователь, не скачавший необходимый контент, либо скачавший пробную версию программы.

Для каждого эксперимента задаются следующие начальные условия:

PIS (ProbIntSoft) — доля контента в Интернете (остальное — твердый контент);

PIL (ProbIllegal) — доля нелегального контента в Интернете;

PUI (ProbUsingIllegal) — вероятность использования нелегального контента (зависит от параметра качества пиратской копии); вероятность использования легального контента принимается равной единице;

PSAI (ProbSearchAgainIllegal) — вероятность повторного поиска после использования нелегального контента в течение некоторого времени;

PSAL (ProbSearchAgainLegal) — вероятность повторного поиска после использования легального контента в течение некоторого времени;

ProbPUsers — доля пользователей, выбирающих контент согласно текущей популярности;

ProbPiratDisks — доступность контента на пиратских дисках;

PiratesPunishRate — доля наказанных пиратов.

Для переходов на диаграмме состояний (рис. 4) используются значения следующих вероятностей:

I_q — вероятность того, что качество пиратской копии удовлетворяет требованию пользователя;

I_c — вероятность того, что бюджетное ограничение позволяет пользователю купить легальный контент;

$$I_{cq} = I_c I_q$$

Правила поведения агентов. Рассмотрим основные действия агентов в модели. Сначала агент (пользователь) выбирает класс контента, затем среди программ выбранного класса он выбирает конкретный программный продукт (по умолчанию равновероятно). Для учета текущей популярности продукта используется параметр *ProbPUsers*. Если *ProbPUsers* > 0, то выбор продукта осуществляется с вероятностью, пропорциональной ее текущей популярности. Популярность программы увеличивается на единицу, как только она скачана или куплена через любой источник. Например, для двух программ из одного класса с популярностями 10 и 30, вероятности выбора равны соответственно 0,25 и 0,75.

После выбора программы для агента устанавливаются значения параметров *BudgetLimit*, *Necessity*, *Quality_required*, *AttemptMax*, *ContactRate*, *Adoption-Fraction* и *MultipleBL*. Реальным бюджетным ограничением для агента является произведение параметров *BudgetLimit* и *Necessity*. Это значение может увеличиться путем умножения на *MultipleBL* в случае, если пользователь получил информацию о наказаниях за пиратство и хочет избежать нарушения закона.

Пользователь с вероятностью *PIS* выбирает Интернет в качестве источника получения контента, или с вероятностью $(1 - PIS)$ будет искать твердый диск с нужной программой. В последнем случае с вероятностью *ProbPiratesDisks* он найдет контрафактный экземпляр со стоимостью *PirateDiskCost*. Тогда выполняется проверка, удовлетворяет ли его этот экземпляр по качеству и стоимости, и если — да, то он покупает пиратскую копию и считается пи-

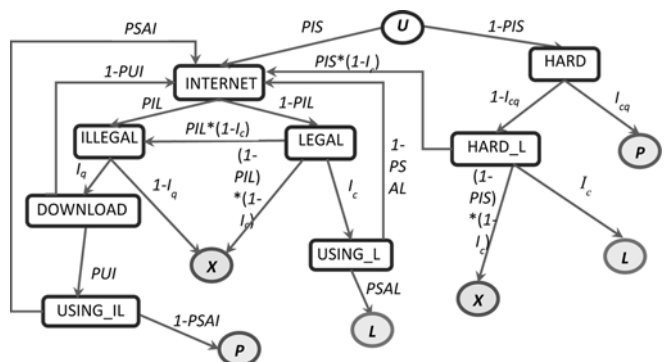


Рис. 4. Диаграмма состояний агентной модели

ратом (P). В противном случае проверяется возможность покупки лицензионного диска, стоимость которого сравнивается с бюджетным ограничением. Если бюджет позволяет, он становится легальным пользователем (L), если — нет, то с вероятностью PIS агент будет искать нужную программу в Интернете или с вероятностью $(1 - PIS)$ прекратит поиск и окажется неудовлетворенным пользователем (X).

Если агент ведет поиск нужной программы в Интернете, то с вероятностью PIL он найдет пиратскую копию программы или с вероятностью $(1 - PIL)$ — легальный экземпляр. В последнем случае, если агенту хватает его бюджетного ограничения, он покупает легальную версию программы и либо прекращает поиск, становясь легальным пользователем с вероятностью $PSAL$, или с вероятностью $(1 - PSAL)$ может продолжить поиск в Интернете (в случае, если его не устроила либо сама программа, либо дополнительные расходы, техподдержка и т. д.). Если бюджетное ограничение не позволяет агенту приобрести легальную версию программы, то с вероятностью PIL он перейдет на сайт с нелегальным контентом или с вероятностью $(1 - PIL)$ перейдет в состояние X , т. е. станет неудовлетворенным пользователем. Заметим, что состояние X означает прекращение поиска только после того, как агент осуществил назначенное ему число попыток поиска $AttemptMax$.

На сайте с нелегальным контентом пользователь скачает пиратскую версию нужной ему программы и, если она удовлетворяет его требованиям к качеству, начнет ею пользоваться, т. е. станет пиратом P , если нет — перейдет в состояние X . Пользователь будет использовать скачанный пиратский дистрибутив с вероятностью PUI или с вероятностью $(1 - PUI)$ продолжит поиск нужной программы на других сайтах. Кроме того, пользователь может возобновить поиск после пробного использования пиратской программы с вероятностью $PSAI$. Если программа его устраивает, он прекращает поиск и становится пиратом P с вероятностью $(1 - PSAI)$.

Вариант модели с обменом информацией о наказаниях за пиратство предусматривает ненулевое значение параметра $PiratesPunishRate > 0$ (доля наказанных пиратов), при этом агенты передают эту информацию своим друзьям (их число — $ContactRate$). Если агент получил сообщение о наказании от своих друзей и с вероятностью $AdaptionFraction$ поверил этой информации, то он не будет скачивать пиратскую версию ни при каких условиях, а будет пытаться приобрести лицензионную программу, если она ему очень нужна (параметр $Necessity \geq 0,7$), увеличив свое бюджетное ограничение в $MultipleBL$ раз. В противном случае агент перейдет в состояние X .

Ключевые статистики. В каждом эксперименте для заданного числа агентов (по умолчанию задается 1000 агентов в эксперименте) вычисляются следующие показатели — ключевые статистики:

- 1) $Pirates$ — число пользователей-пиратов;
- 2) $LegalUsers$ — число легальных пользователей (из них $HardMedia$ — число пользователей, купивших лицензионные твердые диски);
- 3) $Dissatisfied$ — число неудовлетворенных пользователей (которые не стали ни пиратами, ни легальными пользователями, а либо скачали бесплатную пробную версию продукта, либо не приобрели ничего вообще);
- 4) $TotalCost$ — общая выручка от продажи контента;
- 5) $LostProfit$ — упущенная выгода правообладателей (вычисляется на основе предположения, что если необходимость продукта $Necessity > 0,6$, то пользователь бы купил программу);
- 6) $TotalBenefit$ — полезность контента для общества (сумма выгод от всех скачанных и купленных программных продуктов);
- 7) $TotalBenefitNeeded$ — желаемая полезность для общества (сумма выгод от всех программных продуктов, которые необходимы пользователям);
- 8) $LegalSite$ — число посещений легальных сайтов;
- 9) $IllegalSite$ — число посещений нелегальных сайтов;
- 10) $Punished$ — число пользователей, наказанных за пиратство;
- 11) $BlockedPirates$ — число пользователей, которые отказались от пиратства, узнав о наказаниях пиратов от своих друзей;

12) $PartBenefit = \frac{TotalBenefit}{TotalBenefitNeeded}$ — уровень удовлетворенности пользователей;

13) $PartProfit = \frac{TotalCost}{TotalCost + LostProfit}$ — уровень удовлетворенности авторов и правообладателей.

Для того чтобы свести к минимуму случайные отклонения в результатах для каждого фиксированного набора параметров проводилось пять экспериментов. В качестве результатов берутся средние значения статистик.

Значения параметров модели, задаваемые по умолчанию, соответствуют, на наш взгляд, реальным ситуациям выбора пользователями ПО:

- $PIS = 0,8$;
- $PIL = 0,7$;
- $PUI = 0,5$;
- $PSAI = 0,5$;
- $PSAL = 0,2$;
- $ProbPUsers = 0$;
- $ProbPirateDisks = 0,5$;
- $PiratesPunishRate = 0$;
- $CostPirateDisk (CPD) \sim N(400, 200^2)$.

2. Эксперименты на модели

Агентная модель предназначена для экспериментального исследования влияния параметров модели на ключевые статистики. Результаты экс-

периментов могут служить основой для разработки вариантов возможных решений проблемы пиратства в сфере ПО.

Влияние источников доступа к контенту. Очевидно, что возможности доступа к программным продуктам оказывают существенное влияние на поведение потребителей при поиске необходимого ПО и принятии решений о способе его приобретения. Первую группу экспериментов проводили для исследования влияния источников доступа к ПО на вычисляемые результаты путем изменения параметров *PIS* (доля контента в Интернете) и *PIL* (доля нелегального контента в Интернете). Доля твердых пиратских дисков оставалась неизменной. Цены программных продуктов и бюджетные ограничения пользователей были фиксированы. Варьирование обоих параметров выполняли в пределах от 0 до 1 с шагом 0,1. На рис. 5 (см. третью сторону обложки) приведены зависимости числа пиратов (**Pirates**), законопослушных (**LegalUsers**) и неудовлетворенных (**Dissatisfied**) пользователей от доли нелегального контента в Интернете (*PIL*) при изменении доли контента в Интернете (*PIS*).

Естественно, что число пиратов увеличивается с возрастанием доли нелегального контента в Интернете, при этом на рис. 5 легко заметить особое значение $PIL = 40\%$. Если $PIL > 40\%$, то число пиратов увеличивается тем быстрее, чем больше доля контента в Интернете (*PIS*). При $PIL < 40\%$ скорость появления пиратов обратно пропорциональна *PIS*. Уровень удовлетворенности пользователей (**PartBenefit**) растет с увеличением доли нелегального контента в Интернете, при этом можно заметить, что увеличение *PIS* при низких значениях *PIL* снижают уровень удовлетворенности. Однако при высоких значениях *PIL* увеличение *PIS* заметно влияет на уровень удовлетворенности в области значений, не превышающих 0,5, а при дальнейшем увеличении быстро насыщается.

Для легальных пользователей — похожая картина: их доля уменьшается с увеличением *PIL*, причем, чем больше контента в Интернете, тем быстрее. Аналогично уменьшается доля авторов (**PartProfit**), но эффект выражен ярче.

Число неудовлетворенных пользователей уменьшается с увеличением *PIL*, при этом доля контента в Интернете мало влияет на значения этого показателя.

В случае если при выборе контента пользователи ориентируются на текущие популярности продуктов (и выбирают их с вероятностями, пропорциональными текущим популярностям, $ProbPUsers = 1$), общий уровень пиратства при стандартных значениях параметров растет, и, как следствие, все эффекты от изменения параметров становятся более выраженными. То есть доля пиратов быстрее увеличивается, нежели при случайном выборе контента. Это можно объяснить эффектом положительной обратной

связи, т. е. если при начальных условиях пользователи скачивают наиболее доступные пиратские продукты, то следующие пользователи с большей вероятностью выбирают тот контент, который проще скачать (его популярность выше). В итоге доля пиратов еще больше увеличивается, в основном, за счет уменьшения доли неудовлетворенных пользователей.

Цена и качество программных продуктов. Следующую группу экспериментов проводили в целях исследования влияния цены лицензионного ПО и качества нелегальных копий на уровень пиратства. Цены на программные продукты меняли путем установления скидок от 10 до 100 % с шагом 10 % от цены, которая используется по умолчанию.

Уменьшение цен на лицензионное ПО приводит к возрастанию значения I_c , что, в свою очередь, должно сократить число пиратов и увеличить число легальных пользователей. Это вполне логично и интуитивно понятно.

Качество пиратских копий представлено показателем, значения которого заданы экспертными оценками на интервале (0, 2), исходя из предположения, что чем сложнее подделать программный продукт, тем ниже уровень качества пиратской копии. При проведении экспериментов этот показатель изменялся с шагом 0,2.

На рис. 6 (см. четвертую сторону обложки) показано влияние цены легальной копии и качества подделок на число пиратов, легальных и неудовлетворенных пользователей, а также на уровень удовлетворенности потребителей ПО и правообладателей.

Приведенные графики получены при фиксированных значениях $PIS = 0,8$ и $PIL = 0,7$. На графиках можно видеть, что увеличение цены на легальное ПО ведет к увеличению числа пиратов и неудовлетворенных пользователей, и это увеличение тем больше, чем выше качество пиратских копий. Заметим, что даже при нулевой цене легального программного продукта число пиратов не равно нулю, поскольку в этой серии экспериментов 70 % предложения — ПО с пиратских сайтов, выбор которых осуществляется независимо от цены. Если уменьшать долю предложения нелегального контента в Интернете (*PIL*), темп снижения числа пиратов увеличится. Снижение цены приводит к почти пропорциональному увеличению числа легальных и уменьшению числа неудовлетворенных пользователей. При этом качество подделок почти не влияет на число легальных пользователей, но существенно влияет на число неудовлетворенных.

Можно заметить, что при высоком качестве пиратских копий уровень удовлетворенности пользователей — весьма высокий и практически не зависит от цены на программный продукт. При этом уровень доходов авторов и правообладателей стремительно падает с увеличением качества подделок. Снижение цен на легальное ПО также ведет к росту доходов правообладателей, например, при сниже-

нии цен в 2 раза доля доходов растет в 4 раза (сумма доходов в деньгах увеличивается в 2 раза), однако, снижение цен в 4 раза и более становится экономически невыгодным.

Для случая, когда поиск ПО выполняется только в Интернете, причем в условиях отсутствия нелегальных копий ($PIS = 1, PIL = 0$), наиболее выгодным вариантом для обеих сторон (авторов и пользователей) является снижение цены в 4 раза, при этом доля доходов авторов увеличивается в 4,5 раза (в денежном выражении сумма доходов возрастает примерно на 25 %), а доля неудовлетворенных пользователей уменьшается в 8 раз — с 80 до 10 %. Мы также проводили эксперименты для случая, когда в Интернете присутствует только пиратский контент ($PIS = 1, PIL = 1$), чтобы исследовать влияние изменения качества пиратской копии в чистом виде. Результаты этих экспериментов свидетельствуют о том, что улучшение качества подделок приводит сначала к резкому росту пиратства, но после достижения определенного уровня (больше 1,2 на шкале от 0 до 2) почти перестает влиять на число пиратов.

Влияние бюджетных ограничений. Третью серию экспериментов проводили в целях исследования влияния бюджетных ограничений пользователей на уровень пиратства. Бюджетным ограничением мы называем верхнюю границу денежной суммы, которую пользователь готов потратить на приобретение ПО. Можно предположить, что если бы по-

требители не были ограничены в средствах, то пиратов не было бы вообще. В реальности это не так, пиратами становятся не только бедные, поскольку нелегальный контент не только дешевле, но и, как правило, более доступен. Тем не менее, влияние бюджетного ограничения должно быть заметным. Практически все опросы на эту тему демонстрируют важность этого параметра. В данной серии экспериментов все параметры модели, в том числе цены, фиксировались. Бюджетные ограничения варьировали с помощью мультипликатора, значения которого изменяли от 0 до 3 с интервалом 0,3. Для проверки наблюдаемых тенденций была проведена серия экспериментов, в которой значения мультипликатора изменяли от 0 до 10 с шагом 1. Эксперименты проводили при стандартных условиях доступа к контенту ($PIS = 0,8, PIL = 0,7$) и в условиях "чистого Интернета", т. е. когда источником доступа является только Интернет, и там нет нелегального контента ($PIS = 1, PIL = 0$).

Результаты экспериментов приведены на рис. 7. В стандартных условиях ($PIS = 0,8, PIL = 0,7$) увеличение бюджетного ограничения приводит к уменьшению числа пиратов и неудовлетворенных пользователей. Число легальных пользователей и показатели удовлетворенности растут. При увеличении бюджетного ограничения в 3 раза число легальных пользователей и, как следствие, доходы авторов увеличатся в 6 раз, а число пиратов уменьшается

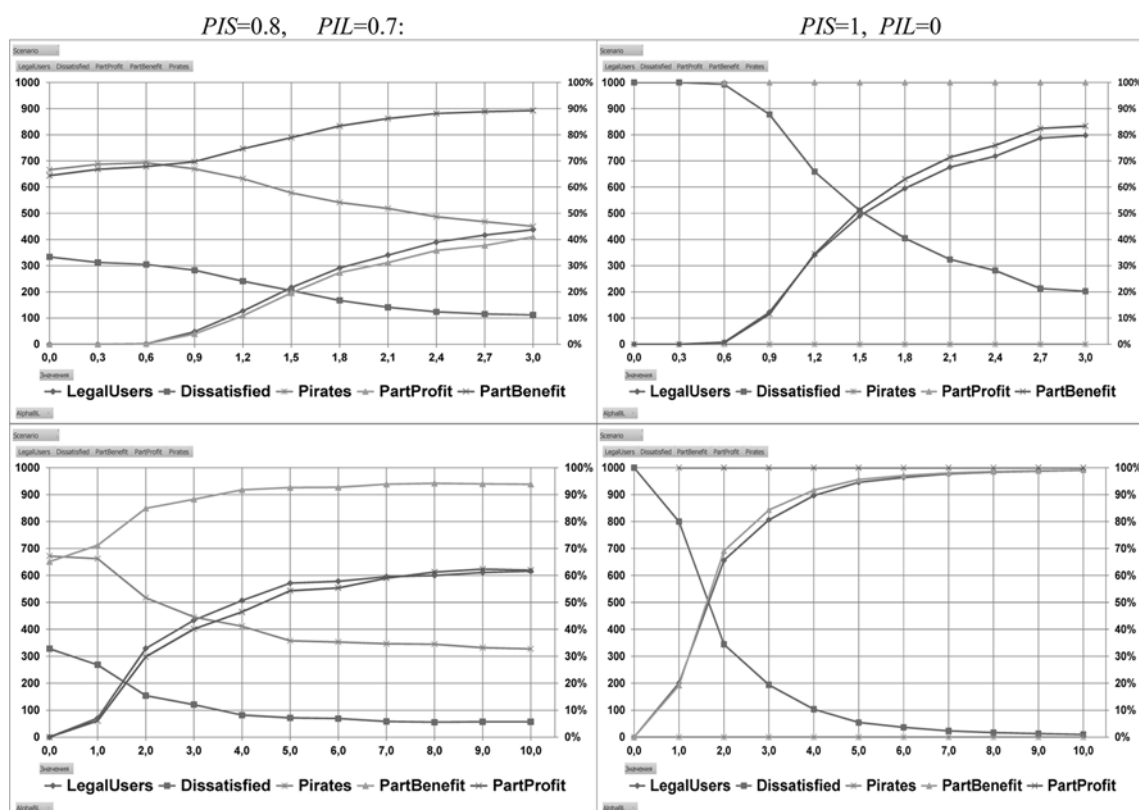


Рис. 7. Влияние бюджетного ограничения на ключевые статистики

в 1,5 раза. При увеличении бюджетного ограничения в 5 и более раз положительные тенденции затухают: число пиратов практически не меняется, а доходы авторов увеличиваются очень слабо (с 55 до 60 %). Даже при 10-кратном увеличении бюджетного ограничения число пиратов превышает 30 %. Это связано с высокой долей нелегального контента в Интернете. При уменьшении этого параметра уровень пиратства снижается более заметно.

В ситуации "чистого Интернета", когда на рынке присутствует только легальный контент, бюджетное ограничение практически определяет уровни удовлетворенности правообладателей и пользователей. Увеличение бюджетного ограничения в 3 раза приводит к 7-кратному росту числа легальных пользователей (и, соответственно, доходов авторов) и уменьшает число неудовлетворенных пользователей в 4 раза. Но после 4-кратного увеличения бюджетного ограничения положительные тенденции затухают.

Обмен информацией о наказании пиратов. В последние годы законодательство в сфере охраны авторских прав в России и в ряде других стран изменяется в направлении ужесточения наказаний нарушителей закона. Уровень пиратства постепенно снижается, и этот факт часто объясняется эффектом запретительных мер. В данной серии экспериментов мы хотели получить ответ на вопрос: как наказания пользователей будут влиять на уровень пиратства? Изменяемым параметром в данном случае была доля пользователей, наказанных за

пиратство (*PiratesPunishRate*), которую варьировали в диапазоне от 0 до 1 с шагом 0,1. Остальные параметры модели в данной серии экспериментов были фиксированы. Результаты экспериментов показаны на рис. 8, где можно заметить, что с увеличением числа наказаний число пиратов снижается, но не становится равным нулю даже тогда, когда все пираты наказаны при значении *PiratesPunishRate* (*PPR*) = 1.

То, что существует ответственность за нарушение закона, известно всем, а вот информация о применении закона менее доступна и более действенна, особенно, если она получена из близкого окружения. Вероятность отказа от пиратских продуктов *B* вычисляли по формуле:

$$B = \frac{PPR \cdot P}{\frac{1}{CR \cdot AF} + PPR}, \quad (1)$$

где *P* — вероятность, с которой пользователь может стать пиратом; *PPR* — *PiratesPunishRate*, *CR* — *ContactRate* — число друзей пользователя, которым он передает информацию о конкретном наказании за пиратство; *AF* — вероятность, с которой пользователь поверит полученной информации. Получается, что вероятность отказа от пиратства пропорциональна доле наказанных пиратов и увеличивается с возрастанием *CR* (число контактов пользователя).

Кроме экспериментов, проведенных в стандартных условиях (значения по умолчанию для всех параметров), мы исследовали модель с выбором про-

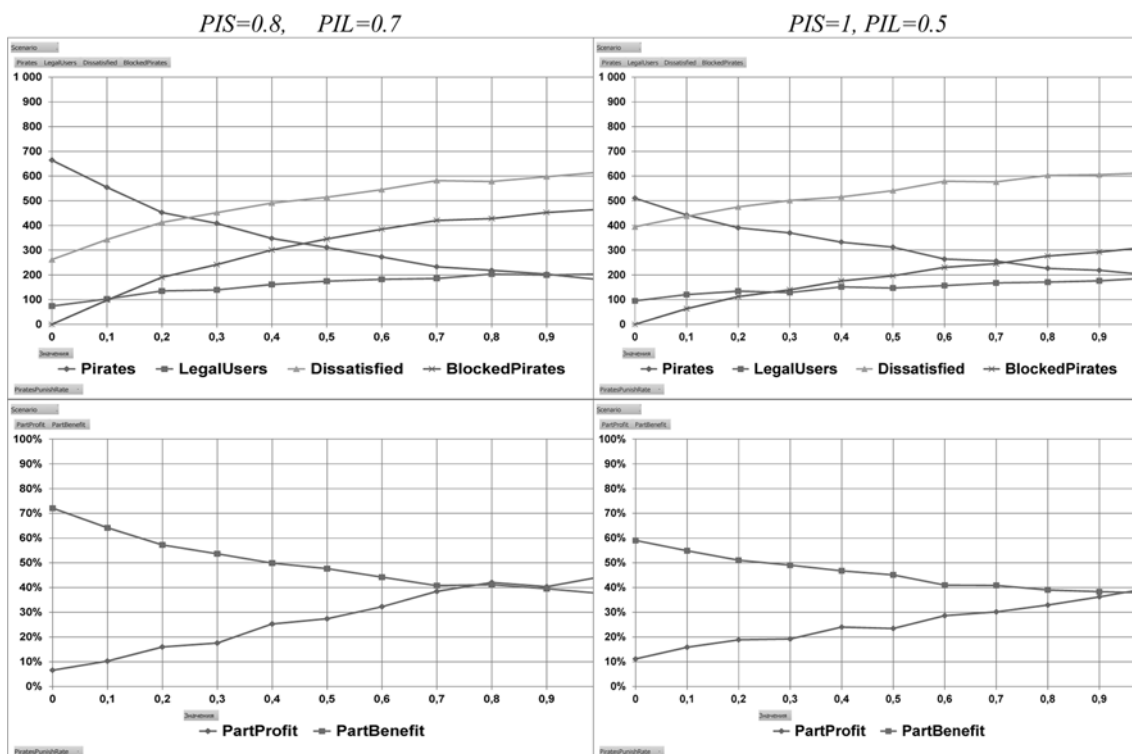


Рис. 8. Изменение ключевых статистик в зависимости от доли наказаний за пиратство

граммных продуктов в соответствии с их текущей популярностью (которая устанавливается после первого эксперимента).

В стандартных условиях ($PIS = 0,8$, $PIL = 0,7$) пользователи, не ставшие пиратами, разделяются на легальных и неудовлетворенных в соотношении примерно 1:3. Следует заметить, что при возрастании доли наказаний от 0 до 1 число легальных пользователей увеличивается незначительно, число неудовлетворенных потребителей возрастает в 2 раза, зато доходы правообладателей становятся существенно больше. Общая удовлетворенность потребителей падает почти в 2 раза, что вполне объяснимо.

Приведенные на рис. 8 графики свидетельствуют о противоположных интересах пользователей и правообладателей ПО. Монотонность зависимостей говорит об отсутствии оптимального значения доли наказаний: с позиций правообладателей наказывать нужно всех, а с позиций потребителей — никого. Однако все вышесказанное справедливо при условии, что потребность пользователей в контенте — достаточно высокая (0,8 из 1,0), а при значениях больше 0,7 они могут увеличивать свои бюджетные ограничения. Результаты экспериментов при значениях $PIS = 1$ и $PIL = 0,5$ демонстрируют те же эффекты, но в меньшей степени, из чего можно заключить, что, чем больше нелегального контента в Интернете, тем сильнее эффект введения наказаний.

Заключение

Результаты моделирования поведения потребителей программных продуктов свидетельствуют о том, что все исследуемые факторы заметно влияют на уровень пиратства и степень удовлетворенности пользователей и правообладателей. Эксперименты с выбором контента согласно его текущей популярности демонстрируют сохранение тенденций и усиление воздействий по сравнению со стандартными условиями. Заметим, что мы изучали пиратство пользователей ПО, подразумевая под этим приобретение ими нелегальных копий. На наш взгляд, более серьезную угрозу представляет другой вид пиратства, обусловленный распространением нелегального ПО через Интернет и на твердых носителях. Пираты этого типа не производят подделок, они просто взламывают оригинальные программы и распространяют их с выгодой для себя. Полученные нами результаты свидетельствуют о том, что в первую очередь следует бороться с пиратами-распространителями, так как самым влиятельным фактором является доля нелегального контента в Интернете. Если она высока, то ни снижение цен, ни увеличение бюджетных ограничений не приводят к существенному снижению числа пиратов. В такой ситуации увеличение доли наказаний дает положительный эффект, проявляющийся в сокращении числа пиратов и росте доходов правообладателей. Однако при этом 75 % потребителей из тех, что

могли бы использовать пиратские программы, останутся неудовлетворенными.

При фиксированном значении доли нелегального контента в Интернете влияние остальных факторов уменьшается в следующем порядке:

1. Увеличение бюджетных ограничений (но не более чем в 4—5 раз) — самый эффективный фактор в отношении роста числа легальных пользователей и общей удовлетворенности потребителей.

2. Снижение цен на лицензионные программные продукты, которое в определенных пределах оказывается выгодным для правообладателей. Чем меньше предложений нелегального контента в Интернете, тем больше снижение цен влияет на уровень пиратства и на уровень общей удовлетворенности потребителей.

3. Усложнение возможностей подделки и взлома легальных продуктов, препятствующее появлению пиратских программ высокого качества. Эксперименты показали, что, чем выше уровень подделки программных продуктов, тем меньше снижение цен влияет на уровень пиратства. Получается, что для сложных продуктов производителям не выгодно снижать цены более чем на 10 %. Но для простых продуктов снижать цены в 2—3 раза оказывается выгодно, поскольку прямым следствием снижения цен является постепенное вытеснение нелегального контента с рынка.

4. Увеличение числа наказаний за пиратство дает положительный эффект, усиливающийся с возрастанием доли нелегального контента в Интернете. При этом доходы правообладателей растут, но общая удовлетворенность потребителей снижается, так как большая часть отказавшихся от пиратства потребителей останутся ни с чем.

Проведенное исследование дает основания полагать, что в сфере распространения программных продуктов самой актуальной задачей является снижение доли нелегального контента в сети Интернет, которая уже сегодня является основным источником распространения ПО. Существующие законы вполне позволяют это сделать, необходимо совершенствовать правоприменительную практику. Кроме того, расширение доступа, снижение цен и усиление защиты от подделок лицензионного ПО способствуют снижению спроса на пиратский контент, следовательно, способствуют уменьшению его предложения.

Заметим, что полученные нами оценки уровня пиратства (65 % в стандартных условиях) очень близки к значениям, приведенным в отчетах BSA (63 %), полученным по другой методике, что можно рассматривать как дополнительный аргумент в пользу их истинности.

Следует помнить о том, что полностью преодолеть пиратство невозможно, и что, кроме издержек, возникающих у производителей и правообладателей, оно предоставляет определенные выгоды

потребителям и обществу в целом. Поэтому практические меры в этой области должны обеспечивать баланс интересов всех субъектов рынка информационных услуг.

Список литературы

1. **Исследование** влияния компьютерного пиратства. Экономические преимущества от снижения уровня компьютерного пиратства / Отчет IDC, BSA, 2010. URL: <http://www.bsa.org>
2. **Software Piracy on the Internet: a threat to your security.** BSA, 2009. URL: <http://www.bsa.org>

3. **Eighth Annual BSA and IDC Global Software Piracy Study,** 2011. URL: <http://globalstudy.bsa.org>
4. **Shadow Market.** 2011 BSA global software piracy study. 2012. URL: <http://www.bsa.org/globalstudy>
5. **Пробелы** в лицензировании: BSA: Глобальное исследование рынка программного обеспечения. 2014. URL: <http://www.bsa.org>
6. **Competitive Advantage: The Economic Impact of Properly Licensed Software.** Global Study Conducted for BSA by INCEAD eL ab. 2013. URL: <http://www.bsa.org>
7. **Андрейчикова О. Н., Андрейчиков А. В.** Что делать с авторским правом в сети Интернет? Поиск решения методом аналитических сетей // Информационные технологии. 2012. № 8. С. 28–35.
8. **Карпов Ю.** Имитационное моделирование систем. Введение в моделирование с AnyLogic5. СПб.: БХВ-Петербург, 2005.

О. Н. Andreichicova¹, Leading Researcher, e-mail: alexandrol@mail.ru,

А. А. Vavilov², Master of Science, **А. Н. Kozyrev**¹, Chief of the Laboratory,

¹Central Economic Mathematical Institute of Russian Academy of Science, Moscow, Russia

²Moscow Institute of Physics and Technology, Moscow, Russia

The Investigation of Consumer Behavior at the Software Market*

The paper describes the research of software consumers' behavior with use of a simulating tool AnyLogic Professional. The research includes the revealing main factors that determine customers' behavior, identification of key decision-making points; a building agent-based model and experiments with it. The outcomes obtained show a good accordance with observed data and can be used for the analysis and a choice of feasible market strategies for software products. The results of simulating software customers' behavior show that all explored factors considerably influence the levels of piracy and satisfaction of users and producers. The experiments with taking into account current popularity of software products demonstrate the same tendencies and an enforcement of impacts in comparison with standard conditions. The estimations of piracy level, which have been obtained in this research (65 % in standard conditions), are very close to the estimates for Russia published in reports of BSA [2] that uses another technique. This fact can be considered as additional proof of their validity.

Keywords: simulation, consumer behavior, software piracy, agent-based model

References

1. **Issledovanie** vlijaniya komp'yuternogo piratstva. Jekonomicheskie preimushhestva ot snizheniya urovnja komp'yuternogo piratstva. *Otchet IDC, BSA*, 2010. URL: <http://www.bsa.org>
2. **Software Piracy on the Internet: a threat to your security.** BSA, 2009. URL: <http://www.bsa.org>
3. **Eighth Annual BSA and IDC Global Software Piracy Study,** 2011. URL: <http://globalstudy.bsa.org>

4. **Shadow Market.** 2011 BSA global software piracy study. 2012. URL: <http://www.bsa.org/globalstudy>
5. **Probeli** v licenzirovanii: BSA: Global'noe issledovanie rynka programmnoho obespechenija 2014. URL: <http://bsa.org>
6. **Competitive Advantage: The Economic Impact of Properly Licensed Software.** *Global Study Conducted for BSA by INCEAD eL ab.* 2013. URL: <http://www.bsa.org>
7. **Andreichikova O. N., Andreichikov A. V.** Chto delat' s avtorskim pravom v seti Internet? Poisk reshenija metodom analiticheskikh setej. *Informacionnye tehnologii.* 2012. N. 8. P. 28–35.
8. **Karpov Ju.** *Imitacionnoe modelirovanie sistem.* Vvedenie v modelirovanie s AnyLogic5. SPb.: BHV-Peterburg, 2005.

*The research is supported by Russian Scientific Foundation, project N 14-18-01999 "Measurement and forecasts in the digital economy".

Адрес редакции:

107076, Москва, Стромынский пер., 4

Телефон редакции журнала (499) 269-5510

E-mail: it@novtex.ru

Технический редактор *Е. В. Конова.*

Корректор *Т. В. Пчелкина.*

Сдано в набор 11.03.2015. Подписано в печать 23.04.2015. Формат 60×88 1/8. Бумага офсетная.

Усл. печ. л. 8,86. Заказ IT515. Цена договорная.

Журнал зарегистрирован в Министерстве Российской Федерации по делам печати, телерадиовещания и средств массовых коммуникаций.

Свидетельство о регистрации ПИ № 77-15565 от 02 июня 2003 г.

Оригинал-макет ООО "Авансед солюшнз". Отпечатано в ООО "Авансед солюшнз".

119071, г. Москва, Ленинский пр-т, д. 19, стр. 1.