

ISSN 2222-8772

**МАТЕМАТИЧЕСКИЕ
СТРУКТУРЫ
И
МОДЕЛИРОВАНИЕ**

**№ 2(46)
2018**



**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМ. Ф.М. ДОСТОЕВСКОГО»**

**МАТЕМАТИЧЕСКИЕ
СТРУКТУРЫ
И
МОДЕЛИРОВАНИЕ**

№ 2(46)

Омск
2018

Математические структуры и моделирование. — Омск : Омский государственный университет, 2018. — № 2(46). — 170 с.
ISSN 2222-8772 (print)
ISSN 2222-8799 (online)

Редакционная коллегия

А. К. Гуц	главный редактор, председатель редакционной коллегии, доктор физ.-мат. наук, профессор, зав. кафедрой кибернетики, Омский государственный университет им. Ф.М. Достоевского.
Д. Н. Лавров	ответственный за выпуск редактор, зам. глав. редактора, канд. техн. наук, доцент, зав. каф. компьютерных технологий и сетей, Омский государственный университет им. Ф.М. Достоевского.
Н. Ф. Богаченко	технический редактор, зам. глав. редактора, канд. физ.-мат. наук, доцент, Омский государственный университет им. Ф.М. Достоевского.
С. В. Белим	д.ф.-м.н., профессор, зав. кафедрой информационной безопасности, проректор по научной работе, Омский государственный университет им. Ф.М. Достоевского.
В. П. Голубятников	доктор физ.-мат. наук, профессор Новосибирского государственного университета, главный научный сотрудник Института математики СО РАН, г. Новосибирск.
С. И. Горлов	доктор физ.-мат. наук, профессор, ректор Нижневартковского государственного университета.
А. Г. Гринь	доктор физ.-мат. наук, профессор, кафедра кибернетики, Омский государственный университет им. Ф.М. Достоевского.
В. А. Еровенко	доктор физ.-мат. наук, профессор, зав. кафедрой общей математики и информатики Белорусского государственного университета, г. Минск, Республика Беларусь.
B. Zilber	Dr.Sc. (Phys.-Math.), Professor of Mathematical Logic, Mathematical Institute, University of Oxford, UK.
А. Н. Кабанов	канд. физ.-мат. наук, кафедра кибернетики, Омский государственный университет им. Ф.М. Достоевского.
А. В. Копыльцов	д.т.н., профессор, кафедра информационных систем, Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина).
А. Г. Коробейников	д.т.н., профессор, зам. директора по науке Санкт-Петербургского филиала Института земного магнетизма, ионосферы и распространения радиоволн им. Н.В. Пушкова РАН.
П. А. Корчагин	доктор техн. наук, профессор, проректор по научной работе, Сибирская государственная автомобильно-дорожная академия (СибАДИ).
V. Kreinovich	Ph.D. (Phys.-Math.), Professor, Computer Science Department, University of Texas at El Paso, Texas, USA.
А. В. Плетюхов	д.ф.-м.н., профессор кафедры общей и теоретической физики Брестского государственного университета им. А.С. Пушкина, Республика Беларусь.
Л. Б. Соколинский	д.ф.-м.н., профессор, проректор по информатизации, зав. кафедрой системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет), г. Челябинск.
A. A. Fedorenko	Ph.D. (Phys.-Math.), Researcher (CR1) at the French National Centre of Scientific Research (CNRS) Laboratoire de Physique de l'ENS-Lyon, France.
A. Jadczyk	Ph.D., Professor, Researcher, Laboratoire de Physique, Universite de Toulouse III et CNRS, France.

Учредитель

Федеральное государственное бюджетное образовательное учреждение высшего образования «Омский государственный университет им. Ф. М. Достоевского».
Свидетельство о регистрации средства массовой информации ПИ № ФС77-72200 от 15 января 2018 г. выдано Роскомнадзором.

Адрес редакции, издателя и типографии

644077, Омская обл., г. Омск,
пр-т Мира, д. 55а.

Дата выхода в свет: 30.06.2018.

Тираж 100 экз.

Свободная цена.

МАТЕМАТИЧЕСКИЕ СТРУКТУРЫ И МОДЕЛИРОВАНИЕ

Журнал основан в 1998 году. В журнале публикуются статьи, в которых излагаются результаты исследований по фундаментальной и прикладной математике, теоретической физике, компьютерным наукам, философии и истории математики и информатики, а также размышления, касающиеся окружающей нас природы и общества. Объекты исследования должны быть представлены в форме некоторых математических структур и моделей.

Все статьи журнала проходят обязательное рецензирование. Рефераты статей журнала опубликованы в «Реферативном журнале» и «Mathematical Reviews» (США). Журнал индексируется в РИНЦ (elibrary.ru) и «Zentralblatt für Mathematik» (Германия). Журнал входит в Перечень рецензируемых научных изданий ВАК РФ, в которых должны быть опубликованы основные результаты диссертаций на соискание учёных степеней (Приказ Минобрнауки России от 25 июля 2014 г. № 793).

Все статьи в журнале публикуются под лицензией Attribution 4.0 International (CC-BY).

Электронная версия журнала представлена в сети:

<http://msm.univer.omsk.su>
<http://msm.omsk.su>

Подписной индекс по каталогу «Пресса России»: 94082

Электронная почта главного редактора:

guts@omsu.ru

Электронная почта выпускающего редактора:

lavrov@omsu.ru

СОДЕРЖАНИЕ

Фундаментальная математика и физика

В.В. Варламов. *Теоретико-групповое описание периодической системы элементов* 5

А.К. Гуц, Г.Б. Гольдина, А.Н. Кабанов. *Аффинные представления как групп Ли преобразований трёхмерных разрешимых групп Ли* 24

Прикладная математика и моделирование

Р.Э. Мамедли. *Устойчивость неоднородных трёхслойных стержней при неравномерном поле температуры в нелинейно упругой среде* .. 33

В.В. Гольдяпин, В.А. Шовин, Е.В. Надей, В.И. Савалкин, Г.В. Нечаева. *Построение дисперсионных комплексов для оценки эффективности иммунотерапии аллергической бронхиальной астмы* 39

В.А. Шовин. *Нелинейные структурные уравнения и квадратичный факторный анализ* 51

А.А. Кондюрина, Д.Н. Лавров. *Результаты эксперимента по обнаружению беспроводной точки доступа модифицированным методом трилатерации* 62

F. Zapata, O. Kosheleva, V. Kreinovich. *Why Under Stress Positive Reinforcement is More Effective? Why Optimists Study Better? Why People Become Restless? Simple Utility-Based Explanations* 66

V. Kreinovich, O. Kosheleva, M. Afravi, G. Bejarano, M. Chacon. *Economics of Commitment: Why Giving Away Some Freedom Makes Sense* .. 73

Компьютерные науки

С.В. Белим, С.Б. Ларионов. *Алгоритм формирования обучающего множества искусственной нейронной сети для сегментации изображения* 79

С.В. Лейхтер, С.Н. Чуканов. *Сравнение изображений на основе построения уравнений Гамильтона* 86

С.В. Гусс, Д.Н. Лавров. *Подходы к реализации сетевого протокола обеспечения гарантированной доставки при мультимаршрутной передаче данных* 95

Продолжение на следующей странице

Наши публикации



O. Kosheleva, V. Kreinovich. *Why Deep Learning Methods Use KL Divergence Instead of Least Squares: A Possible Pedagogical Explanation...* 102

M. Ceberio, V. Kreinovich. *How to Store Tensors in Computer Memory: An Observation* 107

A. Pownuk, V. Kreinovich. *How Interval Measurement Uncertainty Affects the Results of Data Processing: A Calculus-Based Approach to Computing the Range of a Box* 118

Информационная безопасность

Н.И. Синадский, А.В. Агафонов. *Автоматизация тестирования сетевых средств защиты информации на основе применения эволюционно-генетического подхода* 125

Н.Ф. Богаченко. *Анализ проблем управления разграничением доступа в крупномасштабных информационных системах* 135

А.Н. Мироненко. *Метод статического анализа исходного кода приложений операционной системы Android на наличие вредоносного кода* 153

А.Н. Мироненко. *Метод применения (t, n) — пороговой схемы в стеганографии* 157

ТЕОРЕТИКО-ГРУППОВОЕ ОПИСАНИЕ ПЕРИОДИЧЕСКОЙ СИСТЕМЫ ЭЛЕМЕНТОВ

В.В. Варламов

д.ф.-м.н., e-mail: varlamov@subsiiu.ru

Сибирский государственный индустриальный университет, Новокузнецк, Россия

Аннотация. Рассматривается аксиоматическая реализация теоретико-группового описания периодической системы элементов. Периодическая система элементов представляется как единая квантовая система бесструктурных состояний. Вычисляются массы элементов группы суперактиноидов. В рамках алгебраической формулировки единой квантовой системы устанавливается связь с теорией твисторов.

Ключевые слова: периодическая система элементов, единая квантовая система, конформная группа, группа Румера-Фета, твисторная структура.

1. Введение

Согласно модели Бора, атом состоит из положительно заряженного ядра и окружающего его облака отрицательно заряженных электронов. Атомы различных элементов периодической системы отличаются зарядом (числом протонов) их ядер. Атом как целое представляется совокупностью его частей: протонов, нейтронов и электронов. Нетрудно видеть, что в данной модели реализуется классическая редукционистская схема, согласно которой *целое* определяется *частями*, т. е. целое, понимаемое как *агрегат*, является производной от своих частей.

Однако, как известно, *принцип сепарабельности*¹, являющийся исходной посылкой редукционизма, имеет ограниченное применение в квантовой механике. Если подсистемы (части) системы S находятся в *несепарабельном* (запутанном) состоянии, то никакие глобальные свойства системы S , взятой как целое, *не зависят* и *не определяются* свойствами её частей. В рамках запутанной квантовой системы не имеется чистого состояния для какой-либо отдельной подсистемы, т. е. никакая из подсистем $S_1, S_2, \dots, S_N$ не обладает индивидуальным независимым существованием. В любом случае квантовой запутанности недопустимо рассматривать части *квантового целого* как автономные сущности. Иными словами, в случае несепарабельных состояний части

¹Одно из лучших определений этого принципа приводит Каракостас: «Если состояния любых пространственно отделённых подсистем $S_1, S_2, \dots, S_N$ составной системы S точно определены для каждой подсистемы, то состояния составной системы S целиком и полностью определяются состояниями подсистем и их физическими взаимодействиями, включая пространственно-временные связи» [1].

«растворяются» в целом, превращая последнее в бесструктурное состояние. Именно это имел в виду Гейзенберг, говоря, что понятие «состоит из» (принцип сепарабельности) не работает в квантовой области (в физике частиц)². Гейзенберг пишет: «В экспериментах с атомными процессами мы имеем дело с вещами и фактами, которые столь же реальны, сколь реальны любые явления повседневной жизни. Но атомы или элементарные частицы реальны не в такой степени. Они образуют скорее мир тенденций или возможностей, чем мир вещей и фактов» [3, с. 117]. Отсюда следует, что «атомы — не вещи» и что их не следует ставить в один ряд с объектами классической физики, т. е. с макрообъектами, для которых справедлив принцип сепарабельности³. Квантовая система, состояния которой при измерении актуализируются как атомы периодической системы с их «частями» (протонами, нейтронами, электронами), на *потенциальном уровне* существует в суперпозиционном (бесструктурном) состоянии⁴.

Модель Резерфорда–Бора, полностью базирующаяся на предпосылках редукционизма и принципа сепарабельности, имеет своей *холистической анти-тезой* модель Румера–Фета. Уже в первой статье по этой тематике Румер и Фет пишут: «Обычно система элементов объясняется с помощью структурной модели Резерфорда, состоящей из ядра и электронных оболочек, при этом основным параметром, различающим элементы, является атомный заряд. В этой работе атом рассматривается как бесструктурная частица, к которой применяются общие принципы физики симметрии, аналогично тому, как это делается для адронов в SU(6)-классификации. . . Мы называем бесструктурный атом "кулоновской системой"; состояния этой системы должны изображаться векторами пространства, где определено некоторое представление группы **Spin(4)**» [4]. Как отмечает Фет: «Своеобразная черта предлагаемой модели — "игнорирование" структуры атомов, лежащей в основе модели Бора» [5, с. 155]. Другой важнейшей характерной чертой модели Румера–Фета является представление периодической системы элементов как *единой квантовой системы*: «... в то время как модель Бора рассматривает в качестве отдельной квантовой системы атом одного элемента (причём атомный номер входит в теорию в виде параметра, так что имеется столько квантовых систем, сколько элементов), в нашей

² «Как в теории относительности пришлось пожертвовать старым понятием одновременности, а в квантовой механике — понятием электронных орбит, так в физике частиц надо жертвовать понятием деления или понятием «состоит из» [2, с. 77].

³ Модель Бора, являющаяся развитием планетарной модели Резерфорда, по сути ставит атомы в один ряд с макрообъектами классической физики (Солнечная система).

⁴ У Каракостаса находим такие строки: «Вспомним, что с фундаментальной точки зрения квантовой теории физический мир появляется перед нами как неразделённое целое (unbroken whole). Мир сам по себе не является уже разделённым. Мы разделяем его. С целью получить какой-либо тип описания, говорить об объекте, или соотнести с экспериментальными фактами, лежащая в основе природы целостность должна быть разложена на взаимодействующие, но разделённые системы без холистических связей между ними. Это субъектно-объектное разделение иногда метафорически называют «скальпелем Гейзенберга» [1]. Под действием «скальпеля Гейзенберга» целостность рассекается (актуализируется) на «куски» («части»), т. е. целое, существующее в потенции в виде бесструктурного состояния, при измерении распадается на части.

модели атомы всевозможных элементов рассматриваются как состояния единой квантовой системы, соединяемые друг с другом действием группы симметрии» [5, с. 155]. Модель Румера–Фета описывает периодическую систему элементов с позиции холизма, т. е. как *единую квантовую систему бесструктурных состояний*⁵.

В настоящей статье единая квантовая система **U** на потенциальном уровне определяется C^* -алгеброй, состоящей из *оператора энергии* H и присоединённых к H генераторов *группы фундаментальной симметрии* G_f . Состояния системы **U** формируются в рамках конструкции Гельфанда–Наймарка–Сигала (ГНС) [8, 9], т. е. как *циклические представления* операторной алгебры. В силу общности задания системы **U** и гибкости конструкции ГНС при каждой конкретной реализации операторной алгебры (так называемом «одевании» C^* -алгебры) получаем свой (соответствующий данной реализации) спектр состояний системы **U**⁶. В п. 3 настоящей статьи в качестве группы фундаментальной симметрии рассматривается *конформная группа* ($G_f = \text{SO}(2, 4)$). В этом случае конкретная реализация операторной C^* -алгебры задаётся посредством присоединённых к H генераторов комплексной оболочки групповой алгебры $\mathfrak{so}(2, 4)$ и твисторной структуры, ассоциированной с группой $\text{SU}(2, 2)$ (универсальное накрытие конформной группы). Комплексная оболочка алгебры $\mathfrak{so}(2, 4)$ приводит к представлению $F_{ss'}^+$ группы Румера–Фета $\text{SO}(2, 4) \otimes \text{SU}(2) \otimes \text{SU}(2)$, в рамках которой задаётся теоретико-групповое описание периодической системы элементов. В п. 4 вычисляются массы элементов группы суперактиноидов. В п. 5 определяется алгебраическая формулировка единой квантовой системы (реализация операторной алгебры) и устанавливается связь с теорией твисторов (твисторной структурой).

⁵Следует подчеркнуть, что холизм модели Румера–Фета не противостоит редукционизму модели Резерфорда–Бора. Холизм и редукционизм являются двумя взаимодополняющими системами описания *физической реальности*. Согласно концепции Гейзенберга–Фока [2, 6], физическая реальность имеет двухуровневую структуру: *потенциальная реальность* и *актуальная реальность*. Для потенциальной реальности характерно холистическое видение, актуальная реальность — царство редукционизма. Опираясь на теорему Кохена–Спекера и понятие квантовой контекстуальности, де Ронде [7] приходит к выводу, что «физическая реальность $\neq$ актуальная реальность». Описание квантового объекта (атома) в рамках одной только актуальной реальности является *arbitrary* неполным описанием. Перефразируя Гейзенберга, «атомы — более чем вещи».

⁶Так, в случае, когда присоединённые к H генераторы группы фундаментальной симметрии ($G_f = \text{SO}_0(1, 3)$ — группа Лоренца) являются генераторами комплексной оболочки групповой алгебры $\mathfrak{sl}(2, \mathbb{C})$, получаем линейно растущий спектр масс состояний («элементарных частиц») [10, 11]. В этом случае «одевание» операторной алгебры и построение циклических представлений конструкции ГНС осуществляется в рамках спинорной структуры (заряженные, нейтральные, истинно нейтральные (майорановские) состояния и их дискретные симметрии задаются посредством морфизмов спинорной структуры, см. [12–17]). В [11] показано, что массы «элементарных частиц» кратны массе электрона с точностью 0.41 %. Здесь имеет место прямая аналогия с электрическим зарядом. Любой электрический заряд кратен заряду электрона, причём кратен абсолютно точно. Если любой электрический заряд абсолютно кратен (квантован) заряду электрона, то в случае масс эта кратность имеет место с точностью 0.41 % (в среднем).

2. Единая квантовая система

Как уже отмечалось во введении, исходным пунктом построения теоретико-группового описания периодической системы элементов является понятие единой квантовой системы **U**. С целью дать наиболее универсальное определение системы **U** мы должны абстрагироваться от всех макроскопических (локализуемых) определений, принадлежащих исключительно к уровню актуальной реальности (таких как «структура», «частица» и т. д.). Следуя Гейзенбергу, будем считать, что на фундаментальном уровне в основе определения системы **U** лежат два понятия: *энергия* и *симметрия*⁷. Зададим единую квантовую систему **U** посредством следующей системы аксиом (более подробно см. [23], а также [24, 25]).

A.I (Энергия и фундаментальная симметрия) *Единая квантовая система U на фундаментальном уровне характеризуется C^* -алгеброй $\mathfrak{A}$ с единицей, состоящей из оператора энергии H и присоединённых к H генераторов группы фундаментальной симметрии G_f , образующих с H общую систему собственных функций.*

A.II (Состояния) *Физическое состояние C^* -алгебры $\mathfrak{A}$ определяется циклическим вектором $|\Phi\rangle$ представления π C^* -алгебры в сепарабельном гильбертовом пространстве H_∞ :*

$$\omega_\Phi(H) = \frac{\langle \Phi | \pi(H)\Phi \rangle}{\langle \Phi | \Phi \rangle}.$$

Множество $PS(\mathfrak{A})$ всех чистых состояний C^ -алгебры $\mathfrak{A}$ совпадает с множеством всех состояний $\omega_\Phi(H)$, ассоциированных со всеми неприводимыми циклическими представлениями π алгебры $\mathfrak{A}$, $|\Phi\rangle \in H_\infty$ (конструкция Гельфанда–Наймарка–Сигала).*

A.III (Пространство лучей) *Множество всех чистых состояний $\omega_\Phi(H)$ при выполнении условия $\omega_\Phi(H) \geq 0$ образует физическое гильбертово пространство H_{phys} (в общем случае пространство H_{phys} несепарабельно). Для каждого вектора состояния $|\Psi\rangle \in H_{\text{phys}}$ существует единичный луч $\Psi = e^{i\alpha} |\Psi\rangle$, где α пробегает все вещественные числа и $\sqrt{\langle \Psi | \Psi \rangle} = 1$. Пространство лучей есть фактор-пространство $\hat{H} = H_{\text{phys}}/S^1$, т. е. проективное пространство одномерных подпространств из H_{phys} . Все состояния единой квантовой системы **U** описываются единичными лучами.*

A.IV (Аксиома спектральности) *В $\hat{H}$ существует полная система состояний с неотрицательной энергией.*

A.V (Принцип суперпозиции) *Основное соответствие между физическими состояниями и элементами пространства $\hat{H}$ включает принцип суперпозиции квантовой теории, т. е. существует набор базисных состояний*

⁷Что с неизбежностью приводит нас к *неоаристотелеву* взгляду на основания квантовой механики (концепция Гейзенберга–Фока [2, 6]). Следует отметить, что схожих воззрений придерживался Маргенау [18], а также Поппер [19]. Одним из наиболее ярких представителей этого направления является Шимони [20] (см. Каракостас [1], Севальников [21], а также сборник [22] под редакцией Кохена).

таких, что произвольные состояния могут быть построены из них при помощи линейных суперпозиций.

Выберем в качестве фундаментальной симметрии конформную группу. Конформная группа встречается в современной физике в самых различных ситуациях и по сути является столь же универсальной, как группа Лоренца; есть множество релятивистских теорий и, точно так же, конформных⁸.

3. Группа Румера–Фета

После определения конформной группы в качестве фундаментальной симметрии следующим логическим шагом является построение конкретной реализации операторной алгебры. Это построение следует начать с определения комплексной оболочки групповой алгебры $\mathfrak{so}(2, 4)$, что приведёт далее к представлению $F_{ss'}^+$ группы Румера–Фета⁹.

Как известно [5], система пятнадцати генераторов конформной группы $SO(2, 4)$ удовлетворяет следующим перестановочным соотношениям:

$$[\mathbf{L}_{\alpha\beta}, \mathbf{L}_{\gamma\delta}] = i(g_{\alpha\delta}\mathbf{L}_{\beta\gamma} + g_{\beta\gamma}\mathbf{L}_{\alpha\delta} - g_{\alpha\gamma}\mathbf{L}_{\beta\delta} - g_{\beta\delta}\mathbf{L}_{\alpha\gamma}),$$

$$(\alpha, \beta, \gamma, \delta = 1, \dots, 6, \alpha \neq \beta, \gamma \neq \delta).$$

Генераторы $\mathbf{L}_{\alpha\beta}$ образуют базис групповой алгебры $\mathfrak{so}(2, 4)$. С целью перейти к комплексной оболочке алгебры $\mathfrak{so}(2, 4)$ рассмотрим другую систему генераторов, предложенную Т. Йао [30]. Пусть

$$\begin{aligned} \mathbf{J}_1 &= 1/2 (\mathbf{L}_{23} - \mathbf{L}_{14}), & \mathbf{J}_2 &= 1/2 (\mathbf{L}_{31} - \mathbf{L}_{24}), & \mathbf{J}_3 &= 1/2 (\mathbf{L}_{12} - \mathbf{L}_{34}), \\ \mathbf{K}_1 &= 1/2 (\mathbf{L}_{23} + \mathbf{L}_{14}), & \mathbf{K}_2 &= 1/2 (\mathbf{L}_{31} + \mathbf{L}_{24}), & \mathbf{K}_3 &= 1/2 (\mathbf{L}_{12} + \mathbf{L}_{34}), \\ \mathbf{P}_1 &= 1/2 (-\mathbf{L}_{35} - \mathbf{L}_{16}), & \mathbf{P}_2 &= 1/2 (\mathbf{L}_{45} - \mathbf{L}_{36}), & \mathbf{P}_0 &= 1/2 (-\mathbf{L}_{34} - \mathbf{L}_{56}), \\ \mathbf{Q}_1 &= 1/2 (\mathbf{L}_{35} - \mathbf{L}_{46}), & \mathbf{Q}_2 &= 1/2 (\mathbf{L}_{45} + \mathbf{L}_{36}), & \mathbf{Q}_0 &= 1/2 (\mathbf{L}_{34} - \mathbf{L}_{56}), \\ \mathbf{S}_1 &= 1/2 (-\mathbf{L}_{15} + \mathbf{L}_{26}), & \mathbf{S}_2 &= 1/2 (-\mathbf{L}_{25} - \mathbf{L}_{16}), & \mathbf{S}_0 &= 1/2 (\mathbf{L}_{12} - \mathbf{L}_{56}), \end{aligned}$$

⁸Более того, как показал Сигал [26], алгебра Ли неоднородной группы Лоренца (т. е. группы Пуанкаре) может быть получена деформацией из конформной алгебры Ли. В свою очередь конформная алгебра Ли является «жёсткой», т. е. не может быть получена деформированием другой алгебры Ли. В силу этого свойства конформная алгебра (алгебра некомпактной вещественной псевдоортогональной группы в шестимерном пространстве с сигнатурой $(-, -, -, -, +, +)$) имеет уникальный (завершённый) характер и занимает выделенное место среди других алгебр.

⁹Первой работой в этом направлении является [4], где было отмечено поразительное сходство между строением системы химических элементов и строением энергетического спектра атома водорода. Это сходство объясняется в [4] в рамках представления Фока F [27] для группы $\mathbf{Spin}(4)$ (двулистная накрывающая группы $SO(4)$). Однако главным недостатком описания в [4] является приводимость представления F , что не позволяло рассматривать систему как «элементарную» в смысле групповой механики. В 1972 г. Конопельченко [28] расширил представление Фока F до представления F^+ конформной группы, устранив тем самым указанный выше недостаток. Далее, опираясь на связь с нумерацией Маделунга, Фет определяет представления F_s^+ и $F_{ss'}^+$ (для определения представления $F_{ss'}^+$, пришлось изменить «лексикографическое правило» Маделунга). После довольно длительного периода забвения (совершенно незаслуженного!) к модели Румера–Фета возобновился интерес (см. Киблер [29]).

$$\mathbf{T}_1 = 1/2(-\mathbf{L}_{15} - \mathbf{L}_{26}), \quad \mathbf{T}_2 = 1/2(\mathbf{L}_{25} - \mathbf{L}_{16}), \quad \mathbf{T}_0 = 1/2(-\mathbf{L}_{12} - \mathbf{L}_{56}). \quad (1)$$

Эта система 18 генераторов связана тремя соотношениями

$$\mathbf{J}_3 - \mathbf{K}_3 = \mathbf{P}_0 - \mathbf{Q}_0, \quad \mathbf{J}_3 + \mathbf{K}_3 = \mathbf{S}_0 - \mathbf{T}_0, \quad \mathbf{P}_0 + \mathbf{Q}_0 = \mathbf{S}_0 + \mathbf{T}_0. \quad (2)$$

В силу независимости генераторов $\mathbf{L}_{\alpha\beta}$ ($\alpha < \beta$), (1) задаёт избыточную систему генераторов группы $\text{SO}(2, 4)$, из которой можно получить базис алгебры $\mathfrak{so}(2, 4)$, исключив три генератора с помощью (2).

Перейдём к комплексной оболочке алгебры $\mathfrak{so}(2, 4)$, положив

$$\begin{aligned} \mathbf{J}_{\pm} &= \mathbf{J}_1 \pm i\mathbf{J}_2, & \mathbf{P}_{\pm} &= \mathbf{P}_1 \pm i\mathbf{P}_2, & \mathbf{S}_{\pm} &= \mathbf{S}_1 \pm i\mathbf{S}_2, \\ \mathbf{K}_{\pm} &= \mathbf{K}_1 \pm i\mathbf{K}_2, & \mathbf{Q}_{\pm} &= \mathbf{Q}_1 \pm i\mathbf{Q}_2, & \mathbf{T}_{\pm} &= \mathbf{T}_1 \pm i\mathbf{T}_2. \end{aligned} \quad (3)$$

Рассмотрим специальное представление конформной группы $\text{SO}(2, 4)$. Это *локальное* представление группы $\text{SO}(2, 4)$ непосредственно связано с представлением Фока F для подгруппы $\text{SO}(4)$. По сути это представление является расширением представления Фока для $\text{SO}(4)$ до унитарного представления конформной группы $\text{SO}(2, 4)$ в *пространстве Фока* $\mathfrak{F}$ с помощью базиса

$$\begin{aligned} |j, \sigma, \tau\rangle \quad (j = 0, 1/2, 1, 3/2, \dots; \\ \sigma = -j, -j+1, \dots, j-1, j; \tau = -j, -j+1, \dots, j-1, j). \end{aligned} \quad (4)$$

Используя генераторы (3) комплексной оболочке алгебры $\mathfrak{so}(2, 4)$, получим

$$\begin{aligned} \mathbf{J}_- |j, \sigma, \tau\rangle &= \sqrt{(j+\sigma)(j-\sigma+1)} |j, \sigma-1, \tau\rangle, \\ \mathbf{J}_+ |j, \sigma, \tau\rangle &= \sqrt{(j-\sigma)(j+\sigma+1)} |j, \sigma+1, \tau\rangle, \\ \mathbf{J}_3 |j, \sigma, \tau\rangle &= \sigma |j, \sigma, \tau\rangle, \\ \mathbf{K}_- |j, \sigma, \tau\rangle &= \sqrt{(j+\tau)(j-\tau+1)} |j, \sigma, \tau-1\rangle, \\ \mathbf{K}_+ |j, \sigma, \tau\rangle &= \sqrt{(j-\tau)(j+\tau+1)} |j, \sigma, \tau+1\rangle, \\ \mathbf{K}_3 |j, \sigma, \tau\rangle &= \tau |j, \sigma, \tau\rangle, \\ \mathbf{P}_- |j, \sigma, \tau\rangle &= -i\sqrt{(j+\sigma)(j-\tau)} \left| j - \frac{1}{2}, \sigma - \frac{1}{2}, \tau + \frac{1}{2} \right\rangle, \\ \mathbf{P}_+ |j, \sigma, \tau\rangle &= i\sqrt{(j-\tau+1)(j+\sigma+1)} \left| j + \frac{1}{2}, \sigma + \frac{1}{2}, \tau - \frac{1}{2} \right\rangle, \\ \mathbf{P}_0 |j, \sigma, \tau\rangle &= \left(j + \frac{\sigma - \tau + 1}{2} \right) \sigma |j, \sigma, \tau\rangle, \\ \mathbf{Q}_- |j, \sigma, \tau\rangle &= i\sqrt{(j+\tau)(j-\sigma)} \left| j - \frac{1}{2}, \sigma + \frac{1}{2}, \tau - \frac{1}{2} \right\rangle, \\ \mathbf{Q}_+ |j, \sigma, \tau\rangle &= -i\sqrt{(j-\sigma+1)(j+\tau+1)} \left| j + \frac{1}{2}, \sigma - \frac{1}{2}, \tau + \frac{1}{2} \right\rangle, \\ \mathbf{Q}_0 |j, \sigma, \tau\rangle &= \left(j - \frac{\sigma - \tau - 1}{2} \right) |j, \sigma, \tau\rangle, \end{aligned} \quad (5)$$

$$\begin{aligned}
\mathbf{S}_-|j, \sigma, \tau\rangle &= i\sqrt{(j+\sigma)(j+\tau)}\left|j-\frac{1}{2}, \sigma-\frac{1}{2}, \tau-\frac{1}{2}\right\rangle, \\
\mathbf{S}_+|j, \sigma, \tau\rangle &= -i\sqrt{(j+\tau+1)(j+\sigma+1)}\left|j+\frac{1}{2}, \sigma+\frac{1}{2}, \tau+\frac{1}{2}\right\rangle, \\
\mathbf{S}_0|j, \sigma, \tau\rangle &= \left(j+\frac{\sigma+\tau+1}{2}\right)\sigma|j, \sigma, \tau\rangle, \\
\mathbf{T}_-|j, \sigma, \tau\rangle &= -i\sqrt{(j-\tau)(j-\sigma)}\left|j-\frac{1}{2}, \sigma+\frac{1}{2}, \tau+\frac{1}{2}\right\rangle, \\
\mathbf{T}_+|j, \sigma, \tau\rangle &= i\sqrt{(j-\sigma+1)(j-\tau+1)}\left|j+\frac{1}{2}, \sigma-\frac{1}{2}, \tau-\frac{1}{2}\right\rangle, \\
\mathbf{T}_0|j, \sigma, \tau\rangle &= \left(j-\frac{\sigma+\tau-1}{2}\right)|j, \sigma, \tau\rangle.
\end{aligned}$$

Формулы (5) задают унитарное представление конформной группы $\text{SO}(2, 4)$ в пространстве Фока $\mathfrak{F}$. Представление, задаваемое формулами (5), называется *расширением F^+ представления Фока на конформную группу* [5]. Представление F^+ , задаваемое формулами (5), ещё недостаточно для описания периодической системы элементов. С этой целью необходимо включить четвёртое число Маделунга s (аналогичное спину), что приводит к группе (первое «удвоение»)

$$\text{SO}(2, 4) \otimes \text{SU}(2). \quad (6)$$

Представление $F_s^+ = \varphi_2 \otimes F^+$ группы (6), где φ_2 — унитарное представление группы $\text{SU}(2)$ в пространстве $C(2)$, уже удовлетворяет этому требованию (включению числа Маделунга s). Базис пространства $\mathfrak{F}^2 = C(2) \otimes \mathfrak{F}$ представления F_s^+ имеет вид

$$\begin{aligned}
|n, l, m, s\rangle, \quad n = 1, 2, \dots; \quad l = 0, 1, \dots, n-1; \\
m = -l, -l+1, \dots, l-1, l; \quad s = -1/2, 1/2. \quad (7)
\end{aligned}$$

Здесь n, l, m — квантовые числа конформной группы.

Пусть τ_k — генераторы алгебры Ли группы $\text{SU}(2)$, тогда генератор τ_3 перестановочен со всеми генераторами подгруппы $\text{SO}(2, 4) \otimes \mathbf{1}$, а следовательно, и её алгебры Ли, поэтому генераторы $\mathbf{R}_0, \mathbf{L}^2, \mathbf{J}_3 + \mathbf{K}_3, \tau_3$ перестановочны друг с другом. Собственные векторы операторов, представляющих эти генераторы в пространстве $\mathfrak{F}^2$, имеют вид

$$\left|n, l, m, \frac{1}{2}\right\rangle = \begin{bmatrix} \Psi_{nlm}^1 \\ 0 \end{bmatrix}, \quad \left|n, l, m, -\frac{1}{2}\right\rangle = \begin{bmatrix} 0 \\ \Psi_{nlm}^2 \end{bmatrix}$$

с собственными значениями $n, l(l+1), m, \frac{1}{2}$ и $n, l(l+1), m, -\frac{1}{2}$. Действие операторов, представляющих генераторы τ_+, τ_-, τ_3 в пространстве $\mathfrak{F}^2$, задаётся следующими формулами:

$$\tau_+ \left|n, l, m, -\frac{1}{2}\right\rangle = \left|n, l, m, \frac{1}{2}\right\rangle, \quad \tau_- \left|n, l, m, \frac{1}{2}\right\rangle = \left|n, l, m, -\frac{1}{2}\right\rangle,$$

$$\tau_3 |n, l, m, s\rangle = s |n, l, m, s\rangle.$$

Базис $|n, l, m, s\rangle$ находится во взаимно-однозначном соответствии с элементами периодической системы в силу нумерации Маделунга. Связь между расположением элемента в таблице Менделеева и набором чисел (n, l, m, s) задаётся так называемым «лексикографическим правилом» Маделунга $Z \leftrightarrow (n, l, m, s)$ ¹⁰:

- 1) элементы располагаются в порядке возрастания атомного номера Z ;
- 2) наборы (n, l, m, s) располагаются в порядке возрастания $n + l$; при заданном $n + l$, в порядке возрастания n ; при заданных $n + l$, n , в порядке возрастания m ; при заданных $n + l$, n , m , в порядке возрастания s ;
- 3) Z -му элементу ставится в соответствие Z -й набор.

В нумерации Маделунга сумма $n + l$ не имеет группового смысла: это сумма квантового числа n (собственного значения оператора $\mathbf{R}_0 = -\mathbf{L}_{56}$) и числа l , не являющегося квантовым числом (квантовым числом является $l(l + 1)$ — собственное значение оператора $\mathbf{L}^2 = \mathbf{L}_{12}^2 + \mathbf{L}_{23}^2 + \mathbf{L}_{31}^2$, а l всего лишь задаёт это квантовое число). Следовательно, с точки зрения теоретико-группового описания системы элементов число $n + l$ не должно входить в формулировку «лексикографического правила». В [5] Фет вводит новое квантовое число ν , равное $\nu = 1/2(n + l + 1)$ для нечётного значения $n + l$ и $\nu = 1/2(n + l)$ для чётного значения $n + l$. Введение квантового числа ν позволяет изменить нумерацию Маделунга, что приводит к следующему «лексикографическому правилу» (правило Фета):

- 1) элементы располагаются в порядке возрастания атомного номера Z ;
- 2) наборы $(\nu, \lambda, \mu, s, s')$ располагаются в порядке возрастания ν ; при заданном ν — в порядке возрастания s' ; при заданных ν , s' — в порядке убывания λ ; при заданных ν , s' , λ — в порядке возрастания μ ; при заданных ν , s' , λ , μ — в порядке возрастания s ;
- 3) Z -му элементу ставится в соответствие Z -й набор.

Введение пятого квантового числа приводит к ещё одному «удвоению» пространства представления, в котором действует группа Румера–Фета

$$\mathrm{SO}(2, 4) \otimes \mathrm{SU}(2) \otimes \mathrm{SU}(2)'. \quad (8)$$

¹⁰Эрвин Маделунг первым применил «водородные» квантовые числа n, l, m, s для нумерации элементов периодической системы. Следует отметить, что числа n, l, m, s не являются в модели Бора квантовыми числами, поскольку в этой модели нет единого квантово-механического описания системы элементов, последним присваивается атомный номер Z , различающий, а не объединяющий отдельные квантовые системы. Полученную классификацию элементов Маделунг назвал «эмпирической», поскольку он не смог связать её с моделью Бора, что в принципе и не возможно было сделать с позиции редукционизма (известно, что Маделунг показывал свою классификацию Бору). Видимо, именно из-за отсутствия теоретического обоснования на тот момент (20-е годы прошлого столетия) он опубликовал её в виде справочного материала в [31, с. 585]. Теоретическое обоснование (понимание природы чисел Маделунга) было дано впоследствии Фетом [5] с позиции холистического (теоретико-группового) видения. Ещё раз подчеркнём, модели Резерфорда–Бора и Румера–Фета являются не взаимоисключающими, а *взаимодополняющими* друг друга. История чисел Маделунга получила развитие также и в несколько ином направлении в работах Клечковского [32], где заполнение электронных уровней атома рассматривалось согласно правилу последовательного заполнения $(n + l)$ -групп (так называемые группы Маделунга–Клечковского).

Представление $F_{ss'}^+ = \varphi'_2 \otimes F_s^+$ группы (8), где φ'_2 — унитарное представление группы $SU(2)'$ в пространстве $C(2)$, удовлетворяет требованию включения пятого квантового числа. Базис пространства $\mathfrak{F}^4 = C(2) \otimes \mathfrak{F}^2$ представления $F_{ss'}^+$ имеет вид

$$|\nu, \lambda, \mu, s, s'\rangle, \quad \nu = 1, 2, \dots; \lambda = 0, 1, \dots, \nu - 1; \\ \mu = -\lambda, -\lambda + 1, \dots, \lambda - 1, \lambda; s = -1/2, 1/2, s' = -1/2, 1/2. \quad (9)$$

Пусть τ'_k — генераторы алгебры Ли группы $SU(2)'$, тогда генератор τ'_3 перестановочен со всеми генераторами подгруппы $SO(2, 4) \otimes SU(2) \otimes \mathbf{1}$, а следовательно, и её алгебры Ли, поэтому генераторы $\mathbf{R}_0 = -\mathbf{L}_{56} = \mathbf{P}_0 + \mathbf{Q}_0 = \mathbf{S}_0 + \mathbf{T}_0$, $\mathbf{L}^2$, $\mathbf{J}_3 + \mathbf{K}_3$, τ_3 , τ'_3 перестановочны друг с другом. Общие собственные векторы операторов, представляющих эти генераторы в пространстве $\mathfrak{F}^4$, имеют вид

$$\left| \nu, \lambda, \mu, s, +\frac{1}{2} \right\rangle = \begin{bmatrix} \Psi_{\nu\lambda,\mu}^1 \\ \Psi_{\nu\lambda,\mu}^2 \\ 0 \\ 0 \end{bmatrix}, \quad \left| \nu, \lambda, \mu, s, -\frac{1}{2} \right\rangle = \begin{bmatrix} 0 \\ 0 \\ \Psi_{\nu\lambda,\mu}^3 \\ \Psi_{\nu\lambda,\mu}^4 \end{bmatrix}$$

с собственными значениями $\nu, \lambda(\lambda + 1), \mu, s, \frac{1}{2}$ и $\nu, \lambda(\lambda + 1), \mu, s, -\frac{1}{2}$.

Действие операторов, представляющих генераторы τ_+ , τ_- , τ_3 и τ'_+ , τ'_- , τ'_3 в пространстве $\mathfrak{F}^4$, задаётся следующими формулами:

$$\tau_+ \left| \nu, \lambda, \mu, -\frac{1}{2}, s' \right\rangle = \left| \nu, \lambda, \mu, \frac{1}{2}, s' \right\rangle, \quad \tau_- \left| \nu, \lambda, \mu, \frac{1}{2}, s' \right\rangle = \left| \nu, \lambda, \mu, -\frac{1}{2}, s' \right\rangle,$$

$$\tau_3 \left| \nu, \lambda, \mu, s, s' \right\rangle = s \left| \nu, \lambda, \mu, s, s' \right\rangle.$$

$$\tau'_+ \left| \nu, \lambda, \mu, s, -\frac{1}{2} \right\rangle = \left| \nu, \lambda, \mu, s, \frac{1}{2} \right\rangle, \quad \tau'_- \left| \nu, \lambda, \mu, s, \frac{1}{2} \right\rangle = \left| \nu, \lambda, \mu, s, -\frac{1}{2} \right\rangle,$$

$$\tau'_3 \left| \nu, \lambda, \mu, s, s' \right\rangle = s' \left| \nu, \lambda, \mu, s, s' \right\rangle.$$

Адреса элементов (см. рис. 1) задаются набором квантовых чисел группы Румера–Фета (8), нумерующих базисные векторы $|\nu, \lambda, \mu, s, s'\rangle$ пространства $\mathfrak{F}^4$. Генераторы алгебры Ли группы $SO(2, 4)$ действуют на квантовые числа ν, λ, μ по формулам (5) с заменой n, l, m на ν, λ, μ .

		$\nu=1$		$\nu=2$		$\nu=3$		$\nu=4$			
		$s'=-1/2$	$s'=1/2$	$s'=-1/2$	$s'=1/2$	$s'=-1/2$	$s'=1/2$	$s'=-1/2$	$s'=1/2$		
$\lambda = 0$		H He	Li Be	Na Mg	K Ca	Rb Sr	Cs Ba	Fr Ra	Uue Ubn	$s=-1/2$ $s=1/2$	$\mu = 0$
$\lambda = 1$	{			B C	Al Si	Ga Ge	In Sn	Tl Pb	Nh Fl	$s=-1/2$ $s=1/2$	$\mu = -1$
				N O	P S	As Se	Sb Te	Bi Po	Mc Lv	$s=-1/2$ $s=1/2$	$\mu = 0$
				F Ne	Cl Ar	Br Kr	I Xe	At Rn	Ts Og	$s=-1/2$ $s=1/2$	$\mu = 1$
$\lambda = 2$	{					Sc Ti	Y Zr	Lu Hf	Lr Rf	$s=-1/2$ $s=1/2$	$\mu = -2$
						V Cr	Nb Mo	Ta W	Db Sg	$s=-1/2$ $s=1/2$	$\mu = -1$
						Mn Fe	Tc Ru	Re Os	Bh Hs	$s=-1/2$ $s=1/2$	$\mu = 0$
						Co Ni	Rh Pd	Ir Pt	Mt Ds	$s=-1/2$ $s=1/2$	$\mu = 1$
						Cu Zn	Ag Cd	Au Hg	Rg Cn	$s=-1/2$ $s=1/2$	$\mu = 2$
$\lambda = 3$	{							La Ce	Ac Th	$s=-1/2$ $s=1/2$	$\mu = -3$
								Pr Nd	Pa U	$s=-1/2$ $s=1/2$	$\mu = -2$
								Pm Sm	Np Pu	$s=-1/2$ $s=1/2$	$\mu = -1$
								Eu Gd	Am Cm	$s=-1/2$ $s=1/2$	$\mu = 0$
								Tb Dy	Bk Cf	$s=-1/2$ $s=1/2$	$\mu = 1$
								Ho Er	Es Fm	$s=-1/2$ $s=1/2$	$\mu = 2$
								Tm Yb	Md No	$s=-1/2$ $s=1/2$	$\mu = 3$

Рис. 1. Таблица Менделеева в форме основного представления $F_{ss'}^+$ группы Румера–Фета $SO(2, 4) \otimes SU(2) \otimes SU(2)'$

Имеет место следующая цепочка групп:

$$SO(2, 4) \otimes SU(2) \otimes SU(2) \supset SO(4) \otimes SU(2) \supset SO(3) \otimes SU(2),$$

согласно которой осуществляется редукция основного представления $F_{ss'}^+$ группы Румера–Фета по подгруппам цепочки. Так, мультиплеты подгруппы $SO(3) \otimes SU(2)$ изображаются вертикальными прямоугольниками на рис. 1, а заполняющие их элементы составляют известные s -, p -, d - и f -семейства (в частности, лантаноиды и актиноиды выделяются как мультиплеты подгруппы $SO(3) \otimes SU(2)$). Каждый элемент занимает вполне определённое место, заданное его «адресом» в таблице $(\nu, \lambda, \mu, s, s')$, т. е. соответствующими квантовыми числами группы симметрии. Таким образом, атомы всевозможных элементов ставятся во взаимно-однозначное соответствие с векторами ортонормированного базиса (9) пространства основного представления группы Румера–Фета.

4. Массы элементов «острова стабильности»

На рис. 1 жирным шрифтом отмечены элементы, которые ещё не были открыты или не получили своего официального названия при жизни Румера и Фета. Это элементы последнего столбца рис. 1 с квантовыми числами $\nu = 4$ и $s' = 1/2$: **Db** — Дубний, **Sg** — Сиборгий, **Bh** — Борий, **Hs** — Хассий (эка-осмий), **Mt** — Мейтнерий, **Ds** — Дармштадтий, **Rg** — Рентгений, **Cn** — Коперниций (эка-ртуть). Все эти элементы принадлежат мультиплету с квантовым числом $\lambda = 2$. Мультиплет с квантовым числом $\lambda = 1$ ($\nu = 4$, $s' = 1/2$) образуют недавно открытые элементы: **Nh** — Нихоний (эка-таллий), **Fl** — Флеровий (эка-свинец), **Mc** — Московий (эка-висмут), **Lv** — Ливерморий (эка-полоний), **Ts** — Теннессин (эка-астат), **Og** — Оганесон (эка-радон). Мультиплет с квантовым числом $\lambda = 0$ ($\nu = 4$, $s' = 1/2$) образуют ещё не открытые гипотетические элементы **Uue** — Унунений (эка-франций) с предполагаемой атомной массой 316 а.е. и **Ubn** — Унбиний (эка-радий). Все перечисленные элементы завершают заполнение таблицы Менделеева (с 1-го по 120-й номер включительно) до значения квантового числа $\nu = 4$.

Согласно модели Бора, со 121-го номера начинается заполнение g -оболочки (формирование g -семейства). В модели Румера–Фета g -семейство соответствует квантовым числам $\nu = 5$ и $\lambda = 4$ группы симметрии. На сегодняшний день имеется шесть кандидатов (гипотетических элементов) на принадлежность к этому уровню: **Ubu** — Унбиуний, **Ubb** — Унбий, **Ubt** — Унбитрий, **Ubq** — Унбиквадий, **Ubp** — Унбипентий, **Ubh** — Унбигексий. Это элементы группы суперактиноидов так называемого «острова стабильности».

Вычислим массы элементов **Ubb**, ..., **Ubh**. С этой целью воспользуемся массовой формулой¹¹, предложенной в [5]:

$$m = m_0 + a \left[s'(2\nu - 3) - 5\nu + \frac{11}{2} + 2(\nu^2 - 1) \right] - b \cdot \lambda(\lambda + 1), \quad (10)$$

¹¹Эта формула аналогична массовой формуле Гелл–Манна–Окубо для адронов в $SU(3)$ -теории [33, 34], а также формуле Бега–Сингха в $SU(6)$ -теории [35].

где m_0 , a , b — коэффициенты, не выводимые из теории. Формула (10) аналогична «первому возмущению» в SU(3)- и SU(6)-теориях, которое позволяет вычислить среднюю массу элементов мультиплета (аналога «второго возмущения», приводящего к расщеплению масс внутри мультиплета, для группы Румера–Фета нет). В таб. 1 приведены средние массы «тяжёлых» мультиплетов ($\nu = 3, 4$) периодической системы элементов, вычисленных согласно формуле (10) при $m_0 = 1$, $a = 17$, $b = 5, 5$. Из таб. 1 видно, что точность между экспериментальными и теоретическими значениями масс возрастает по мере увеличения «тяжести» мультиплета, следовательно, формула (10) является асимптотической. Исключение составляет последний мультиплет ($\nu = 4, s' = 1/2, \lambda = 0$), состоящий из гипотетических элементов **Uue** и **Ubn**, массы которых не являются экспериментально подтверждёнными.

Табл. 1. Средние массы «тяжёлых» мультиплетов

	Мультиплет	Масса (эксп.)	Масса (теор.)	Погр-сть, %
1.	$(\nu = 3, s' = -1/2, \lambda = 2)$	55,31	53	−4,36
2.	$(\nu = 3, s' = -1/2, \lambda = 1)$	76,65	75	−2,2
3.	$(\nu = 3, s' = -1/2, \lambda = 0)$	86,54	86	−0,63
4.	$(\nu = 3, s' = 1/2, \lambda = 2)$	99,76	104	4,07
5.	$(\nu = 3, s' = 1/2, \lambda = 1)$	123,51	126	1,98
6.	$(\nu = 3, s' = 1/2, \lambda = 0)$	135,12	137	1,37
7.	$(\nu = 4, s' = -1/2, \lambda = 3)$	154,59	156	0,90
8.	$(\nu = 4, s' = -1/2, \lambda = 2)$	187,96	189	0,55
9.	$(\nu = 4, s' = -1/2, \lambda = 1)$	210,21	211	0,37
10.	$(\nu = 4, s' = -1/2, \lambda = 0)$	224,52	222	−1,13
11.	$(\nu = 4, s' = 1/2, \lambda = 3)$	244,56	241	−1,48
12.	$(\nu = 4, s' = 1/2, \lambda = 2)$	273,10	274	0,33
13.	$(\nu = 4, s' = 1/2, \lambda = 1)$	290,83	296	1,75
14.	$(\nu = 4, s' = 1/2, \lambda = 0)$	318*	307	−3,58*

Найдём среднюю массу первого мультиплета группы суперактиноидов ($\nu = 5, s' = -1/2, \lambda = 4$). Полагая $m_0 = 1$, $a = 17$, $b = 5, 5$, из формулы (10) получим $m = 316$. Средняя масса второго мультиплета суперактиноидов ($\nu = 5, s' = 1/2, \lambda = 4$) при тех же значениях констант равна $m = 435$. Экстраполируя разность средних масс между теоретическими и экспериментальными значениями для «тяжёлых» мультиплетов ($\nu = 3, 4, 5$), получим приближенные значения масс элементов группы суперактиноидов (см. таб. 2). Эти элементы являются частью *расширенной периодической системы* (таблицы Сиборга¹²). В таб. 2 верхний левый индекс обозначает атомный номер элемента, а правый нижний — его массу.

¹²Теоретико-групповое описание таблицы Сиборга в рамках группы Румера–Фета выходит за рамки настоящей статьи и будет подробно рассмотрено в следующей работе.

Табл. 2. Группа суперактиноидов

		$(\nu = 5, s' = -1/2, \lambda = 4)$	$(\nu = 5, s' = 1/2, \lambda = 4)$
$\mu = -4$	$s = -1/2$	$^{121}\mathbf{Ubu}_{308}$	$^{171}\mathbf{Usu}_{423}$
	$s = 1/2$	$^{122}\mathbf{Ubb}_{311}$	$^{172}\mathbf{Usb}_{426}$
$\mu = -3$	$s = -1/2$	$^{123}\mathbf{Ubt}_{313}$	$^{173}\mathbf{Ust}_{428}$
	$s = 1/2$	$^{124}\mathbf{Ubq}_{316}$	$^{174}\mathbf{Usq}_{431}$
$\mu = -2$	$s = -1/2$	$^{125}\mathbf{Ubp}_{318}$	$^{175}\mathbf{Usp}_{433}$
	$s = 1/2$	$^{126}\mathbf{Ubh}_{321}$	$^{176}\mathbf{Ush}_{436}$
$\mu = -1$	$s = -1/2$	$^{127}\mathbf{Ubs}_{322}$	$^{177}\mathbf{Uss}_{437}$
	$s = 1/2$	$^{128}\mathbf{Ubo}_{325}$	$^{178}\mathbf{Uso}_{440}$
$\mu = 0$	$s = -1/2$	$^{129}\mathbf{Ube}_{327}$	$^{179}\mathbf{Use}_{442}$
	$s = 1/2$	$^{130}\mathbf{Utn}_{330}$	$^{180}\mathbf{Uon}_{445}$
$\mu = 1$	$s = -1/2$	$^{131}\mathbf{Utu}_{331}$	$^{181}\mathbf{Uou}_{446}$
	$s = 1/2$	$^{132}\mathbf{Utb}_{334}$	$^{182}\mathbf{Uob}_{449}$
$\mu = 2$	$s = -1/2$	$^{133}\mathbf{Utt}_{336}$	$^{183}\mathbf{Uot}_{451}$
	$s = 1/2$	$^{134}\mathbf{Utg}_{339}$	$^{184}\mathbf{Uog}_{454}$
$\mu = 3$	$s = -1/2$	$^{135}\mathbf{Utp}_{340}$	$^{185}\mathbf{Uop}_{455}$
	$s = 1/2$	$^{136}\mathbf{Uth}_{343}$	$^{186}\mathbf{Uoh}_{458}$
$\mu = 4$	$s = -1/2$	$^{137}\mathbf{Uts}_{345}$	$^{187}\mathbf{Uos}_{460}$
	$s = 1/2$	$^{138}\mathbf{Uto}_{348}$	$^{188}\mathbf{Uoo}_{463}$

5. Реализация операторной алгебры

Как известно, в основании алгебраической формулировки квантовой теории лежит конструкция Гельфанда–Найма–Сигала (ГНС), заключающаяся в каноническом соответствии $\omega \leftrightarrow \pi_\omega$ между состояниями и циклическими представлениями C^* -алгебры [36–38].

Пусть, согласно аксиомы **A.I** (п. 2), к оператору энергии H присоединены генераторы конформной группы, являющейся в данном контексте группой фундаментальной симметрии ($G_f = \text{SO}(2, 4)$). Следовательно, каждое собственное подпространство $\mathbf{H}_E$ оператора энергии инвариантно относительно операторов представления F^+ конформной группы¹³. Это позволяет получить конкретную реализацию («одевание») операторной алгебры $\pi(\mathfrak{A}) \rightarrow \pi(H)$, где $\pi \equiv F^+$. Таким образом, каждое возможное значение энергии (энергетический уровень) является векторным состоянием вида (аксиома **A.II**):

$$\omega_\Phi(H) = \frac{\langle \Phi | \pi(H)\Phi \rangle}{\langle \Phi | \Phi \rangle} = \frac{\langle \Phi | F^+(H)\Phi \rangle}{\langle \Phi | \Phi \rangle},$$

где $|\Phi\rangle$ — циклический вектор гильбертова пространства $\mathbf{H}_\infty$.

¹³Это следует из сходимости комплексных оболочек групповых алгебр $\mathfrak{sl}(2, \mathbb{C})$, $\mathfrak{so}(4)$ и $\mathfrak{so}(2, 4)$.

Далее, в силу изоморфизма $SU(2, 2) \simeq \mathbf{Spin}_+(2, 4)^{14}$, будем рассматривать универсальную накрывающую $\tilde{G}_f$ как *спинорную группу*, что позволит дополнительно ассоциировать с каждым циклическим вектором $|\Phi\rangle \in H_\infty$ *твисторную структуру*. Спинтензорные представления группы $\tilde{G}_f = \mathbf{Spin}_+(2, 4)$ образуют субстрат конечномерных представлений $\tau_{k/2, r/2}, \bar{\tau}_{k/2, r/2}$ конформной группы, реализуемых в пространствах $\text{Sym}_{(k, r)} \subset \mathbb{S}_{2k+r}$ и $\bar{\text{Sym}}_{(k, r)} \subset \bar{\mathbb{S}}_{2k+r}$, где $\mathbb{S}_{2k+r}$ — спинпространство. Действительно, твистор $\mathbf{Z}^\alpha = (s^\alpha, s_{\dot{\alpha}})$ является вектором фундаментального представления группы $\mathbf{Spin}_+(2, 4)$, где $\alpha, \dot{\alpha} = 0, 1$. Вектором общего спинтензорного представления группы $\mathbf{Spin}_+(2, 4)$ является

$$\mathbf{Z} = \begin{bmatrix} \mathbf{S} \\ \bar{\mathbf{S}} \end{bmatrix}, \quad (11)$$

где $\mathbf{S}$ — спинтензор вида

$$\mathbf{S} = s_{\dot{\alpha}_1 \dot{\alpha}_2 \dots \dot{\alpha}_r}^{\alpha_1 \alpha_2 \dots \alpha_k} = \sum s^{\alpha_1} \otimes s^{\alpha_2} \otimes \dots \otimes s^{\alpha_k} \otimes s_{\dot{\alpha}_1} \otimes s_{\dot{\alpha}_2} \otimes \dots \otimes s_{\dot{\alpha}_r}, \quad \alpha_i, \dot{\alpha}_i = 0, 1;$$

т. е. вектор спинпространства $\mathbb{S}_{2k+r} = \mathbb{S}_{2k} \otimes \dot{\mathbb{S}}_{2r}$, где $\dot{\mathbb{S}}_{2r}$ — дуальное спинпространство. $\bar{\mathbf{S}}$ — спинтензор из сопряжённого пространства $\bar{\mathbb{S}}_{2k+r}$. Симметризируя каждый из спинтензоров $\mathbf{S}$ и $\bar{\mathbf{S}}$, входящих в (11), получим симметрический *твисттензор* $\mathbf{Z}$. В свою очередь, как известно [39], спинпространство является минимальным левым идеалом алгебры Клиффорда $\mathcal{C}_{p,q}$, т. е. существует изоморфизм $\mathbb{S}_{2m}(\mathbb{K}) \simeq I_{p,q} = \mathcal{C}_{p,q}f$, где f — примитивный идемпотент алгебры $\mathcal{C}_{p,q}$, $\mathbb{K} = f\mathcal{C}_{p,q}f$ — кольцо деления для $\mathcal{C}_{p,q}$, $m = (p+q)/2$. Комплексное спинпространство $\mathbb{S}_{2m}(\mathbb{C})$ является комплексификацией $\mathbb{C} \otimes I_{p,q}$ минимального левого идеала $I_{p,q}$ вещественной подалгебры $\mathcal{C}_{p,q}$. Следовательно, $\mathbb{S}_{2k+r}$ является минимальным левым идеалом комплексной алгебры $\mathbb{C}_{2k} \otimes \mathbb{C}_{2r} \simeq \mathbb{C}_{2(k+r)}$ (более подробно см. [41, 42]).

Определим систему *базисных циклических векторов*, наделённых комплексной твисторной структурой¹⁵, и соответствующих системе конечномерных

¹⁴Изоморфизм

$$SU(2, 2) = \left\{ \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathbb{C}_4 : \det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = 1 \right\} \simeq \mathbf{Spin}_+(2, 4)$$

следует из алгебраического определения группы Клиффорда–Липшица $\mathbf{F}_{p,q}$ (см. [39, 40]):

$$\mathbf{Spin}_+(2, 4) = \left\{ s \in \begin{bmatrix} \mathbb{C} \otimes \mathcal{C}_{1,1}^0 - i\mathbb{C} \otimes \mathcal{C}_{1,1}^3 & -\mathbb{C} \otimes \mathcal{C}_{1,1}^1 + i\mathbb{C} \otimes \mathcal{C}_{1,1}^2 \\ \mathbb{C} \otimes \mathcal{C}_{1,1}^1 + i\mathbb{C} \otimes \mathcal{C}_{1,1}^2 & \mathbb{C} \otimes \mathcal{C}_{1,1}^0 + i\mathbb{C} \otimes \mathcal{C}_{1,1}^3 \end{bmatrix} \middle| N(s) = 1 \right\},$$

где $\mathbb{C}_4$ — алгебра Дирака, $\mathcal{C}_{1,1}$ — алгебра анти-кватернионов.

¹⁵По сути твисторная структура есть «удвоение» спинорной структуры. В связи с этим возникает вопрос: какой физический смысл имеет спинорная (твисторная) структура, ассоциируемая в данном контексте с циклическими векторами гильбертова пространства? Ответ на этот вопрос может дать хорошо известная аналогия между спинором и кубитом. Как известно, *кубит* есть несепарабельное состояние, заданное суперпозицией $|\psi\rangle = a|0\rangle + b|1\rangle$, где $a, b \in \mathbb{C}$. Кубит есть минимально возможный (элементарный) вектор состояния. Любой вектор состояния может

представлений конформной группы:

$$\begin{aligned}
& | \mathbb{C}_0, \tau_{0,0}(H)\Phi \rangle; \\
& | \mathbb{C}_2, \tau_{1/2,0}(H)\Phi \rangle, \quad | \dot{\mathbb{C}}_2, \tau_{0,1/2}(H)\Phi \rangle; \\
& | \mathbb{C}_2 \otimes \mathbb{C}_2, \tau_{1,0}(H)\Phi \rangle, \quad | \mathbb{C}_2 \otimes \dot{\mathbb{C}}_2, \tau_{1/2,1/2}(H)\Phi \rangle, \quad | \dot{\mathbb{C}}_2 \otimes \dot{\mathbb{C}}_2, \tau_{0,1}(H)\Phi \rangle; \\
& | \mathbb{C}_2 \otimes \mathbb{C}_2 \otimes \mathbb{C}_2, \tau_{3/2,0}(H)\Phi \rangle, \quad | \mathbb{C}_2 \otimes \mathbb{C}_2 \otimes \dot{\mathbb{C}}_2, \tau_{1,1/2}(H)\Phi \rangle, \\
& \quad | \mathbb{C}_2 \otimes \dot{\mathbb{C}}_2 \otimes \dot{\mathbb{C}}_2, \tau_{1/2,1}(H)\Phi \rangle, \quad | \dot{\mathbb{C}}_2 \otimes \dot{\mathbb{C}}_2 \otimes \dot{\mathbb{C}}_2, \tau_{0,3/2}(H)\Phi \rangle; \\
& \dots\dots\dots \\
& | \overline{\mathbb{C}_0, \tau_{0,0}(H)\Phi} \rangle; \\
& | \overline{\mathbb{C}_2, \tau_{1/2,0}(H)\Phi} \rangle, \quad | \overline{\dot{\mathbb{C}}_2, \tau_{0,1/2}(H)\Phi} \rangle; \\
& | \overline{\mathbb{C}_2 \otimes \mathbb{C}_2, \tau_{1,0}(H)\Phi} \rangle, \quad | \overline{\mathbb{C}_2 \otimes \dot{\mathbb{C}}_2, \tau_{1/2,1/2}(H)\Phi} \rangle, \quad | \overline{\dot{\mathbb{C}}_2 \otimes \dot{\mathbb{C}}_2, \tau_{0,1}(H)\Phi} \rangle; \\
& | \overline{\mathbb{C}_2 \otimes \mathbb{C}_2 \otimes \mathbb{C}_2, \tau_{3/2,0}(H)\Phi} \rangle, \quad | \overline{\mathbb{C}_2 \otimes \mathbb{C}_2 \otimes \dot{\mathbb{C}}_2, \tau_{1,1/2}(H)\Phi} \rangle, \\
& \quad | \overline{\mathbb{C}_2 \otimes \dot{\mathbb{C}}_2 \otimes \dot{\mathbb{C}}_2, \tau_{1/2,1}(H)\Phi} \rangle, \quad | \overline{\dot{\mathbb{C}}_2 \otimes \dot{\mathbb{C}}_2 \otimes \dot{\mathbb{C}}_2, \tau_{0,3/2}(H)\Phi} \rangle. \\
& \dots\dots\dots
\end{aligned}$$

Следовательно, согласно конструкции ГНС (аксиома **A.II**), имеем комплексные векторные состояния вида

$$\begin{aligned}
\omega_{\Phi}^c(H) &= \frac{\langle \Phi | \mathbb{C}_{2(k+r)}, \tau_{k/2, r/2}(H)\Phi \rangle}{\langle \Phi | \Phi \rangle}, \\
\bar{\omega}_{\Phi}^c(H) &= \frac{\langle \Phi | \overline{\mathbb{C}_{2(k+r)}, \tau_{k/2, r/2}(H)\Phi} \rangle}{\langle \Phi | \Phi \rangle}.
\end{aligned}$$

Согласно (11), пары $(\omega_{\Phi}^c(H), \bar{\omega}_{\Phi}^c(H))$ образуют *нейтральные состояния*. Далее, множество всех чистых состояний $(\omega_{\Phi}^c(H), \bar{\omega}_{\Phi}^c(H))$ при выполнении условия $\omega_{\Phi}^c(H) \geq 0$ ($\bar{\omega}_{\Phi}^c(H) \geq 0$) образует физическое гильбертово пространство $\mathbf{H}_{\text{phys}}$ (аксиома **A.III**) и, соответственно, пространство лучей $\hat{H} = \mathbf{H}_{\text{phys}}/S^1$. Все состояния физической квантовой системы $\mathbf{U}$ описываются единичными лучами

быть представлен множеством таких элементарных векторов, по этой причине кубит является исходным «строительным кирпичом» для всех остальных векторов состояния любой размерности. Квантовое состояние N кубитов может быть описано вектором гильбертова пространства размерности 2^N . Очевидно, что это пространство совпадает со спинпространством $\mathbb{S}_{2^N}$. Выберем в качестве ортонормированного базиса для этого пространства состояния, в которых каждый кубит имеет определённое значение $|0\rangle$ или $|1\rangle$. Тогда векторы базиса могут быть представлены двоичными строками вида $|x\rangle = |01110010 \dots 1001\rangle$. Общий нормированный вектор может быть выражен в этом базисе как $\sum_{x=0}^{2^N-1} a_x |x\rangle$, где a_x — комплексные числа, удовлетворяющие условию $\sum_x |a_x|^2 = 1$. Уже здесь видна глубокая аналогия между кубитами и 2-компонентными спинорами. Подобно кубитам 2-компонентные спиноры являются «строительными кирпичами» спинорной структуры (посредством тензорных произведений алгебр $\mathbb{C}_2$ и $\dot{\mathbb{C}}_2$). Это означает, что в некотором смысле спинорная (твисторная) структура может быть отождествлена с информационной структурой (на что ранее уже указывал Пенроуз), прообразами которой являются все материальные объекты (в духе «it from bit» Уилера или «it from qubit» д'Ариано).

и при данной реализации операторной алгебры соответствуют атомам периодической системы элементов. При этом имеет место принцип суперпозиции (аксиома **A.V**).

Следуя классификации Гейзенберга [2], всё множество групп симметрий G следует разбить на два класса: 1) *группы фундаментальных (первичных) симметрий* G_f , участвующие в образовании векторов состояния единой квантовой системы $\mathbf{U}$; 2) *группы динамических (вторичных) симметрий* G_d , описывающие приближенные симметрии между векторами состояния системы $\mathbf{U}$. Динамические симметрии G_d связывают между собой различные состояния квантовой системы $\mathbf{U}$ (векторы состояния $|\Psi\rangle \in \mathbf{H}_{\text{phys}}$). Симметрия G_d системы $\mathbf{U}$ может быть представлена как *квантовый переход* между её состояниями (уровнями спектра состояний системы $\mathbf{U}$).

Покажем, что группа Румера–Фета является группой динамической симметрии. Действительно, группа (8) эквивалентна $\widetilde{\text{SO}}(2, 4) \otimes \text{SU}(2) = \text{SU}(2, 2) \otimes \text{SU}(2)$ (см. [43]), поскольку одно «удвоение» в (8) уже фактически учитывается двуделистной накрывающей $\text{SU}(2, 2)$ конформной группы¹⁶. При этом атомы различных элементов ставятся во взаимно-однозначное соответствие с векторами базиса (9) пространства представления F_{ss}^+ . Здесь имеет место прямая аналогия с физикой «элементарных частиц». Согласно Вигнеру [46], квантовая система,

¹⁶На протяжении всей статьи термин «удвоение» встречался много раз. «Удвоение» (или «двуделение и уменьшение симметрии» Паули) является одним из ведущих принципов теоретико-группового (холистического) описания. В письме к Гейзенбергу Паули пишет: «Двуделение и уменьшение симметрии, вот где зарыт Фаустов пудель. Двуделение — старый атрибут дьявола (недаром сомнение называют “раздвоенностью”)» [2, с. 288]. Гейзенберг следующим образом комментирует письмо Паули: «В прежней физике оболочки атома ещё можно было опираться на наглядные образы, заимствованные из репертуара классической физики. Принцип соответствия Бора фиксировал как раз пусть ограниченную, но применимость подобных образов. Однако уже в том, что касается оболочки атома, математическое описание происходящих в ней процессов значительно превосходило эти образы по степени своей абстрактности. Можно было даже соотносить с одним и тем же реальным положением вещей два различных и противоречащих друг другу образа, а именно корпускулярное и волновое представления. В физике же элементарных частиц эти образы уже, по существу, совсем непригодны. Эта физика ещё более абстрактна. Для формулировки природных законов здесь не остаётся поэтому никакой иной отправной точки, кроме свойств симметрии, воплощённых в природе, или, выражаясь иначе, преобразований симметрии (например, смещений или поворотов), которые изначально организуют пространство природы. Но тогда мы неизбежно приходим к вопросу о том, почему существуют именно такие, а не иные преобразования симметрии. Процесс раздвоения, или двуделения, как я его себе представляю, мог бы нам здесь многое объяснить, потому что он каким-то очень естественным образом расширяет пространство природы, создавая тем самым возможность новых симметрий. В идеальном случае можно было бы думать, что все реальные симметрии возникли как следствие подобных раздвоений» [3, с. 342-343]. Под «уменьшением симметрии» следует понимать групповую редукцию, т. е. если имеется цепочка вложенных друг в друга групп $G \supset G_1 \supset G_2 \supset \dots \supset G_k$ и дано неприводимое унитарное представление $\mathfrak{P}$ группы G в пространстве $\mathbf{H}_{\text{phys}}$, то редукция G/G_1 представления $\mathfrak{P}$ группы G по подгруппе G_1 приводит к разложению $\mathfrak{P}$ в ортогональную сумму неприводимых представлений $\mathfrak{P}_i^{(1)}$ подгруппы G_1 . В свою очередь, редукция G_1/G_2 представления группы G_1 по подгруппе G_2 приводит к разложению представлений $\mathfrak{P}_i^{(1)}$ на неприводимые представления $\mathfrak{P}_{ij}^{(2)}$ группы G_2 и т. д. (см. [44, 45]). Таким образом происходит редукция («уменьшение симметрии» Паули) группы G с высокой симметрией на более низкие симметрии подгрупп.

описываемая неприводимым унитарным *представлением* группы Пуанкаре $\mathcal{P}$, называется элементарной частицей. С другой стороны, в согласии с $SU(3)$ - и $SU(6)$ -теориями, элементарная частица описывается *вектором* неприводимого представления группы $SU(3)$ (или $SU(6)$). Следовательно, имеем две теоретико-групповые интерпретации элементарной частицы: как *представления* группы $\mathcal{P}$ (группы фундаментальной симметрии) и как *вектора* представления группы динамической симметрии $SU(3)$ ($SU(6)$). Кроме того, структура массовой формулы (10) для группы Румера–Фета аналогична массовым формулам Гелл–Манна–Окубо и Бега–Сингха для групп $SU(3)$ и $SU(6)$. Действие группы $G_d = SU(2, 2) \otimes SU(2)$, поднятое в $\mathbf{H}_{\text{phys}}$ посредством центрального расширения, переводит друг в друга векторы состояния $|\Psi\rangle \in \mathbf{H}_{\text{phys}}$, соответствующие различным атомам периодической системы элементов.

ЛИТЕРАТУРА

1. Karakostas V. Forms of Quantum Nonseparability and Related Philosophical Consequences // Journal for General Philosophy of Science. 2004. V. 35. P. 283–312.
2. Гейзенберг В. Шаги за горизонт. М. : Прогресс, 1987.
3. Гейзенберг В. Физика и философия. Часть и целое. М. : Наука, 1990.
4. Румер Ю.Б., Фет А.И. Группа $\text{Spin}(4)$ и таблица Менделеева // ТМФ. 1971. Т. 9. С. 203–209.
5. Фет А.И. Группа симметрии химических элементов. Новосибирск : Наука, 2010.
6. Фок В.А. Об интерпретации квантовой механики // УФН. 1957. Т. 62. С. 461–474.
7. de Ronde C. Quantum Superpositions and the Representation of Physical Reality Beyond Measurement Outcomes and Mathematical Structures // arXiv:1603.06112 [quant-ph] (2016).
8. Gelfand I., Neumark M. On the Imbedding of Normed Rings into the Ring of Operators in Hilbert Space // Rec. Math. [Mat. Sbornik] N.S. 1943. V. 12(54). P. 197–217.
9. Segal I. Postulates for general quantum mechanics // Ann. Math. 1947. V. 48. P. 930–948.
10. Варламов В.В. Спектр материи Гейзенберга в абстрактно-алгебраическом подходе // Математические структуры и моделирование. 2016. № 3(39). С. 5–23.
11. Варламов В.В. Квантование массы и группа Лоренца // Математические структуры и моделирование. 2017. № 2(42). С. 11–28.
12. Varlamov V.V. Fundamental Automorphisms of Clifford Algebras and an Extension of Dąbrowski Pin Groups // Hadronic J. 1999. V. 22. P. 497–535.
13. Varlamov V.V. Discrete Symmetries and Clifford Algebras // Int. J. Theor. Phys. 2001. V. 40. P. 769–805.
14. Варламов В.В. Дискретные симметрии на пространствах фактор-представлений группы Лоренца // Математические структуры и моделирование. 2001. Вып. 7. С. 114–127.
15. Varlamov V.V. CPT groups for spinor field in de Sitter space // Phys. Lett. B. 2005. V. 631. P. 187–191.
16. Varlamov V.V. CPT Groups of Higher Spin Fields // Int. J. Theor. Phys. 2012. V. 51. P. 1453–1481.

17. Varlamov V.V. CPT groups of spinor fields in de Sitter and anti-de Sitter spaces // Adv. Appl. Clifford Algebras. 2015. V. 25. P. 487–516.
18. Margenau H. The Nature of physical reality. New York : McGraw Hill, 1950.
19. Popper K. A world of propensities. Bristol : Thoemmes Press, 1995.
20. Shimony A. Search for a naturalistic world view. Vol. 2. Natural science and metaphysics. Cambridge : Cambridge University Press, 1993.
21. Севальников А.Ю. Интерпретации квантовой механики: в поисках новой онтологии. М. : Книжный дом «ЛИБРОКОМ», 2009.
22. Cohen R.S., Horne M., Stachel J. Potentiality, Entanglement and Passion-at-a-Distance. Dordrecht : Springer, 1997.
23. Варламов В.В. О системе аксиом нелокальной квантовой теории // Математические структуры и моделирование. 2017. № 4(44). С. 5–25.
24. Варламов В.В. Комплексный момент и спин-зарядовое гильбертово пространство // Математические структуры и моделирование. 2015. № 4(36). С. 5–22.
25. Varlamov V.V. Spinor Structure and Matter Spectrum // Int. J. Theor. Phys. 2016. V. 55. P. 5008–5045.
26. Segal I. A class of operator algebras which are determined by groups // Duke Math. J. 1951. V. 18. P. 221–265.
27. Фок В.А. Атом водорода и неевклидова геометрия // Изв. АН СССР. Сер. VII. Отд-ние мат. и естеств. наук. 1935. № 2. С. 169–179.
28. Конопельченко Б.Г. Группа $SO(2, 4) + R$ и таблица Менделеева // СО РАН СССР. Ин-т ядерной физики. Препр. Новосибирск, 1972.
29. Kibler M.R. From the Mendeleev periodic table to particle physics and back to periodic table // Foundations of Chemistry. 2006. V. 9. P. 221–234.
30. Yao Tsu. Unitary Irreducible Representations of $SU(2, 2)$ // J. Math. Phys. 1967. V. 8. P. 1931–1954.
31. Маделунг Э. Математический аппарат физики. М. : Наука, 1968.
32. Клечковский В.М. Распределение атомных электронов и правило последовательного заполнения $(n + l)$ -групп. М. : Атомиздат, 1968.
33. Gell-Mann M. Symmetries of Baryons and Mesons // Phys. Rev. 1962. V. 125. P. 1067–1084.
34. Okubo S., Ryan C. Quadratic mass formula in $SU(3)$. Nuovo Cimento. 1964. V. 34. P. 776–779.
35. Bég M., Singh V. Splitting of the 70-Plet of $SU(6)$. Phys. Rev. Lett. 1964. V. 13. P. 509–511.
36. Эмх Ж. Алгебраические методы в статистической механике и квантовой теории поля. М. : Мир, 1976.
37. Боголюбов Н.Н., Логунов А.А., Оксак А.И., Тодоров И.Т. Общие принципы квантовой теории поля. М. : Наука, 1987.
38. Хоружий С.С. Введение в алгебраическую квантовую теорию поля. М. : Наука, 1986.
39. Lounesto P. Clifford Algebras and Spinors. Cambridge: Cambridge Univ. Press, 2001.
40. Varlamov V.V. Universal Coverings of Orthogonal Groups // Adv. Appl. Clifford Algebras. 2004. V. 14. P. 81–168;
41. Varlamov V.V. Generalized Weierstrass representation for surfaces in terms of Dirac-Hestenes spinor field // J. Geometry and Physics. 2000. V. 32. P. 241–251.

42. Варламов В.В. Спинорная структура и периодичность алгебр Клиффорда // Математические структуры и моделирование. 2015. № 3(35). С. 4–20.
43. Фет А.И. Конформная симметрия химических элементов // ТМФ. 1975. Т. 22. С. 323–334.
44. Варламов В.В. Спинорная структура и SU(3)-симметрия // Математические структуры и моделирование. 2015. № 1(33). С. 18–33.
45. Varlamov V.V. Spinor Structure and Internal Symmetries // Int. J. Theor. Phys. 2015. V. 54. P. 3533–3576.
46. Wigner E.P. On unitary representations of the inhomogeneous Lorentz group // Ann. Math. 1939. V. 40. P. 149–204.

GROUP THEORETICAL DESCRIPTION OF PERIODIC SYSTEM OF ELEMENTS

V.V. Varlamov

Dr.Sc. (Phys.-Math.), e-mail: varlamov@sibsiu.ru

Siberian State Industrial University, Novokuznetsk, Russia

Abstract. An axiomatic realization of the group theoretical description is considered for periodic system of elements. The periodic system is represented as a single quantum system of structureless states. Masses of elements of the superactinide group are calculated. A relationship between an algebraic formulation of the single quantum system and twistor theory is established.

Keywords: Mendelev periodic system, single quantum system, conformal group, Rumer-Fet group, twistor structure.

Дата поступления в редакцию: 12.05.2018

АФФИННЫЕ ПРЕДСТАВЛЕНИЯ КАК ГРУПП ЛИ ПРЕОБРАЗОВАНИЙ ТРЁХМЕРНЫХ РАЗРЕШИМЫХ ГРУПП ЛИ

А.К. Гуц

д.ф.-м.н., профессор, e-mail: guts@omsu.ru

Г.Б. Гольдина

студент, e-mail: zoloto13@list.ru

А.Н. Кабанов

к.ф.-м.н., доцент, e-mail: kabanovan@omsu.ru

Омский государственный университет им. Ф.М. Достоевского, Омск, Россия

Аннотация. С помощью метода Ямагучи дается представление как групп Ли аффинных преобразований всех трехмерных разрешимых групп Ли.

Ключевые слова: Аффинная структура, аффинное представление, трёхмерные разрешимые группы Ли.

Введение

В различных исследованиях группы Ли используются, будучи представленными в форме групп аффинных преобразований в аффинном пространстве. Имеются различные способы построения таких представлений. В этой статье строится аффинное представление 3-мерных разрешимых групп Ли методом Ямагучи [1, 2].

1. Аффинные представления групп Ли

Пусть G – вещественная связная группа Ли. Она допускает *аффинное представление*, если существует гомоморфизм

$$\alpha : G \rightarrow \text{Aff}(\mathbb{R}^n) \cong GL(\mathbb{R}^n),$$

$$G \ni g \mapsto \alpha(g) \in \text{Aff}(\mathbb{R}^n),$$

где $\text{Aff}(\mathbb{R}^n)$ – группа аффинных преобразований n -мерного арифметического пространства $\mathbb{R}^n$, или, равно, общих линейных преобразований пространства $\mathbb{R}^n$.

2. Метод Ямагучи

Пусть дана n -мерная группа Ли G_n , её алгебра Ли $\mathfrak{g}_n$ с базисом $X_1, \dots, X_n$. Строим представление

$$\Lambda : \mathfrak{g}_n \rightarrow \mathfrak{gl}(\mathfrak{g}_n),$$

$$\Lambda(X_i) : \mathfrak{g}_n \ni X_j \rightarrow \Lambda(X_i)X_j \in \mathfrak{g}_n$$

такое, что

$$\Lambda(X_i)X_j - \Lambda(X_j)X_i = [X_i, X_j]. \quad (1)$$

Рассматриваем алгебру Ли группы Ли $\text{Aff}(\mathbb{R}^n)$:

$$\mathfrak{aff}(\mathbb{R}^n) = \left\{ \left(\begin{array}{c|c} A & a \\ \hline 0 & 0 \end{array} \right) : A - n \times n - \text{матрица}, a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^n \right\}.$$

Определяем представление

$$\rho : \mathfrak{g}_n \rightarrow \mathfrak{aff}(\mathbb{R}^n),$$

полагая

$$\rho(X) = \left(\begin{array}{c|c} \Lambda(X) & \begin{matrix} t_1 \\ \vdots \\ t_n \end{matrix} \\ \hline 0 & 0 \end{array} \right), \quad X = \sum_{i=1}^n t_i X_i \in \mathfrak{g}_n,$$

$$\rho(X_i) = \left(\begin{array}{c|c} \Lambda(X_i) & e_i \\ \hline 0 & 0 \end{array} \right), \quad e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} i.$$

Пусть $G(n)$ – аналитическая подгруппа группы $\text{Aff}(\mathbb{R}^n)$ с алгеброй Ли $\rho(\mathfrak{g}_n)$. Тогда $G(n) \cong G_n$ и $G(n)$ действует просто транзитивно на $\mathbb{R}^n = \{(x^1, \dots, x^n, 1)^{tr} : x^i \in \mathbb{R}\}$ (см. [1, 2]).

Имеем 1-параметрическую подгруппу $\exp(t_i X_i)$ группы Ли G_n , которой соответствует аффинное преобразование $\exp(t_i \rho(X))$: в $\mathbb{R}^n = \{(x^1, \dots, x^n, 1) : x^i \in \mathbb{R}\}$:

$$G_n \ni \exp(t_i X_i) \rightarrow \exp(t_i \rho(X)) = e^{t_i \rho(X_i)} \in G(n),$$

$$e^{t_i \rho(X_i)} : \begin{pmatrix} x^1 \\ \vdots \\ x^n \\ \hline 1 \end{pmatrix} \rightarrow e^{t_i \rho(X_i)} \begin{pmatrix} x^1 \\ \vdots \\ x^n \\ \hline 1 \end{pmatrix}, \quad t_i \in \mathbb{R}. \quad (2)$$

Преобразования (2) порождают аффинное представление $\alpha(G_n)$:

$$\alpha : G_n \rightarrow GL(\mathfrak{g}_n) \cong GL(\mathbb{R}^n),$$

$$\alpha(\exp(t_i \Lambda(X_i))) \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} = e^{t_i \Lambda(X_i)} \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} + t_i e_i, \quad t_i \in \mathbb{R},$$

$$\exp : \mathfrak{gl}(\mathfrak{g}_n) \rightarrow GL(\mathfrak{g}_n)$$

группы Ли G_n с параметрами $(t_1, \dots, t_n)$, действующее просто транзитивно на $\mathbb{R}^3$, если, конечно, можно найти матрицы $\Lambda(X_i)$, удовлетворяющие условиям (1) [1].

3. Построение аффинных представлений 3-мерных групп Ли

Теорема 1. Любая связная 3-мерная разрешимая вещественная группа Ли G_3 имеет аффинное представление $\alpha(G_3)$, действующее просто транзитивно на $\mathbb{R}^3$.

Доказательство. Используем классификацию 3-мерных вещественных разрешимых алгебр Ли из [3].

1. $C_3 I$.

$$[X_i, X_j] = 0.$$

Здесь $\Lambda(X_i) = 0$, поэтому

$$\rho(X_i) = \left(\begin{array}{c|c} 0 & e_i \\ \hline 0 & 0 \end{array} \right), \quad \exp(t_i \rho(X_i)) = \left(\begin{array}{c|c} E & t_i e_i \\ \hline 0 & 1 \end{array} \right), \quad E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

и соответствующее аффинное преобразование

$$\begin{aligned} (x^1, x^2, x^3) &\rightarrow (x^1 + t_1, x^2, x^3), \\ (x^1, x^2, x^3) &\rightarrow (x^1, x^2 + t_2, x^3), \\ (x^1, x^2, x^3) &\rightarrow (x^1, x^2, x^3 + t_3). \end{aligned} \tag{3}$$

Следовательно, имеем аффинное представление абелевой группы $G_3 I$:

$$(x^1, x^2, x^3) \rightarrow (x^1 + t_1, x^2 + t_2, x^3 + t_3).$$

2. $C_3 II$.

$$\begin{aligned} [X_1, X_2] &= 0, \quad [X_1, X_3] = 0, \quad [X_2, X_3] = X_1, \\ \Lambda(X_1) &= \Lambda(X_2) = 0, \quad \Lambda(X_3) = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

$$X_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Тогда имеем

$$e^{t_1 \rho(X_1)} = \left(\begin{array}{ccc|c} 1 & 0 & 0 & t_1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline & & & 1 \end{array} \right), \quad e^{t_2 \rho(X_2)} = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & t_2 \\ 0 & 0 & 1 & 0 \\ \hline & & & 1 \end{array} \right),$$

$$e^{t_3 \rho(X_3)} = \left(\begin{array}{ccc|c} 1 & -t_3 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & t_3 \\ \hline & & & 1 \end{array} \right)$$

и соответствующее аффинное преобразование

$$\begin{aligned} (x^1, x^2, x^3) &\rightarrow (x^1 + t_1, x^2, x^3), \\ (x^1, x^2, x^3) &\rightarrow (x^1, x^2 + t_2, x^3), \\ (x^1, x^2, x^3) &\rightarrow (x^1 - t_3 x_2, x^2, x^3 + t_3). \end{aligned} \tag{4}$$

Таким образом, группа Ли $G_3 II$ имеет следующее аффинное представление:

$$\begin{cases} \bar{x}^1 = x^1 - t_3 x^2 + t_1, \\ \bar{x}^2 = x^2 + t_2, \\ \bar{x}^3 = x^3 + t_3. \end{cases} \tag{5}$$

3. $C_3 III$.

$$[X_1, X_2] = 0, \quad [X_2, X_3] = 0, \quad [X_3, X_1] = -X_1,$$

$$\Lambda(X_1) = \Lambda(X_2) = 0, \quad \Lambda(X_3) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Тогда имеем

$$e^{t_1 \rho(X_1)} = \left(\begin{array}{ccc|c} 1 & 0 & 0 & t_1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline & & & 1 \end{array} \right), \quad e^{t_2 \rho(X_2)} = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & t_2 \\ 0 & 0 & 1 & 0 \\ \hline & & & 1 \end{array} \right),$$

$$e^{t_3\rho(X_3)} = \left(\begin{array}{ccc|c} e^{-t_3} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & t_3 \\ \hline 0 & & & 1 \end{array} \right)$$

и соответствующее аффинное преобразование

$$\begin{aligned} (x^1, x^2, x^3) &\rightarrow (x^1 + t_1, x^2, x^3), \\ (x^1, x^2, x^3) &\rightarrow (x^1, x^2 + t_2, x^3), \\ (x^1, x^2, x^3) &\rightarrow (e^{-t_3}x^1, x^2, x^3 + t_3). \end{aligned} \quad (6)$$

Таким образом, группа Ли G_3III имеет следующее аффинное представление:

$$\begin{cases} \bar{x}^1 = e^{-t_3}x^1 + t_1, \\ \bar{x}^2 = x^2 + t_2, \\ \bar{x}^3 = x^3 + t_3. \end{cases} \quad (7)$$

4. C_3IV .

$$[X_1, X_2] = 0, \quad [X_2, X_3] = X_1 + X_2, \quad [X_3, X_1] = -X_1,$$

$$\Lambda(X_1) = \Lambda(X_2) = 0, \quad \Lambda(X_3) = \begin{pmatrix} -1 & -1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Тогда имеем

$$\begin{aligned} e^{t_1\rho(X_1)} &= \left(\begin{array}{ccc|c} 1 & 0 & 0 & t_1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & & & 1 \end{array} \right), \quad e^{t_2\rho(X_2)} = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & t_2 \\ 0 & 0 & 1 & 0 \\ \hline 0 & & & 1 \end{array} \right), \\ e^{t_3\rho(X_3)} &= \left(\begin{array}{ccc|c} \sum_{n=0}^{\infty} \frac{(-1)^n t_3^n}{n!} & \sum_{n=1}^{\infty} \frac{(-1)^n t_3^n}{(n-1)!} & 0 & 0 \\ 0 & \sum_{n=0}^{\infty} \frac{(-1)^n t_3^n}{n!} & 0 & 0 \\ 0 & 0 & 1 & t_3 \\ \hline 0 & & & 1 \end{array} \right) \end{aligned}$$

и соответствующее аффинное преобразование

$$\begin{aligned} (x^1, x^2, x^3) &\rightarrow (x^1 + t_1, x^2, x^3), \\ (x^1, x^2, x^3) &\rightarrow (x^1, x^2 + t_2, x^3), \\ (x^1, x^2, x^3) &\rightarrow (e^{-t_3}x^1 + t_3e^{-t_3}x_2, e^{-t_3}x^2, x^3 + t_3). \end{aligned} \quad (8)$$

Таким образом, группа Ли G_3IV имеет следующее аффинное представление:

$$\begin{cases} \bar{x}^1 = e^{-t_3}x^1 + t_3e^{-t_3}x^2 + t_1, \\ \bar{x}^2 = e^{-t_3}x^2 + t_2, \\ \bar{x}^3 = x^3 + t_3. \end{cases} \quad (9)$$

5. C_3V .

$$[X_1, X_2] = 0, \quad [X_2, X_3] = X_2, \quad [X_3, X_1] = -X_1, \\ \Lambda(X_1) = \Lambda(X_2) = 0, \quad \Lambda(X_3) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Тогда имеем

$$e^{t_1\rho(X_1)} \left(\begin{array}{ccc|c} 1 & 0 & 0 & t_1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & & & 1 \end{array} \right), \quad e^{t_2\rho(X_2)} = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & t_2 \\ 0 & 0 & 1 & 0 \\ \hline 0 & & & 1 \end{array} \right), \\ e^{t_3\rho(X_3)} = \left(\begin{array}{ccc|c} e^{-t_3} & 0 & 0 & 0 \\ 0 & e^{-t_3} & 0 & 0 \\ 0 & 0 & 1 & t_3 \\ \hline 0 & & & 1 \end{array} \right)$$

и соответствующее аффинное преобразование

$$\begin{aligned} (x^1, x^2, x^3) &\rightarrow (x^1 + t_1, x^2, x^3), \\ (x^1, x^2, x^3) &\rightarrow (x^1, x^2 + t_2, x^3), \\ (x^1, x^2, x^3) &\rightarrow (e^{-t_3}x^1, e^{-t_3}x^2, x^3 + t_3). \end{aligned} \quad (10)$$

Таким образом, группа Ли G_3V имеет следующее аффинное представление:

$$\begin{cases} \bar{x}^1 = e^{-t_3}x^1 + t_1, \\ \bar{x}^2 = e^{-t_3}x^2 + t_2, \\ \bar{x}^3 = x^3 + t_3. \end{cases} \quad (11)$$

6. C_3VI .

$$[X_1, X_2] = 0, \quad [X_2, X_3] = qX_2 \quad (q \neq 0, 1), \quad [X_3, X_1] = -X_1, \\ \Lambda(X_1) = \Lambda(X_2) = 0, \quad \Lambda(X_3) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -q & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Тогда имеем

$$e^{t_1\rho(X_1)} = \left(\begin{array}{ccc|c} 1 & 0 & 0 & t_1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline & 0 & & 1 \end{array} \right), \quad e^{t_2\rho(X_2)} = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & t_2 \\ 0 & 0 & 1 & 0 \\ \hline & 0 & & 1 \end{array} \right),$$

$$e^{t_3\rho(X_3)} = \left(\begin{array}{ccc|c} e^{-t_3} & 0 & 0 & 0 \\ 0 & e^{-qt_3} & 0 & 0 \\ 0 & 0 & 1 & t_3 \\ \hline & 0 & & 1 \end{array} \right)$$

и соответствующее аффинное преобразование

$$\begin{aligned} (x^1, x^2, x^3) &\rightarrow (x^1 + t_1, x^2, x^3), \\ (x^1, x^2, x^3) &\rightarrow (x^1, x^2 + t_2, x^3), \\ (x^1, x^2, x^3) &\rightarrow (e^{-t_3}x^1, e^{-qt_3}x^2, x^3 + t_3). \end{aligned} \tag{12}$$

Таким образом, группа Ли G_3VI имеет следующее аффинное представление:

$$\begin{cases} \bar{x}^1 = e^{-t_3}x^1 + t_1, \\ \bar{x}^2 = e^{-qt_3}x^2 + t_2, \\ \bar{x}^3 = x^3 + t_3. \end{cases} \tag{13}$$

7. C_3VII .

$$[X_1, X_2] = 0, \quad [X_2, X_3] = -X_1 + qX_2 \quad (q^2 < 1), \quad [X_3, X_1] = -X_2,$$

$$\Lambda(X_1) = \Lambda(X_2) = 0, \quad \Lambda(X_3) = \left(\begin{array}{ccc} 0 & 1 & 0 \\ -1 & -q & 0 \\ 0 & 0 & 0 \end{array} \right).$$

Тогда имеем

$$e^{t_1\rho(X_1)} = \left(\begin{array}{ccc|c} 1 & 0 & 0 & t_1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline & 0 & & 1 \end{array} \right), \quad e^{t_2\rho(X_2)} = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & t_2 \\ 0 & 0 & 1 & 0 \\ \hline & 0 & & 1 \end{array} \right),$$

$$e^{t_3\rho(X_3)} = \left(\begin{array}{ccc|c} e^{-\frac{t_3q}{2}}A(t_3) & 2e^{-\frac{t_3q}{2}}B(t_3) & 0 & 0 \\ -2e^{-\frac{t_3q}{2}}B(t_3) & e^{-\frac{t_3q}{2}}C(t_3) & 0 & 0 \\ 0 & 0 & 1 & t_3 \\ \hline & 0 & & 1 \end{array} \right),$$

где

$$\begin{aligned} A(t) &= \cos \frac{t(4-q^2)}{2\sqrt{|q^2-4|}} + \frac{q}{\sqrt{|q^2-4|}} \sin \frac{t(4-q^2)}{2\sqrt{|q^4-4|}}, \\ B(t) &= \frac{1}{\sqrt{|q^2-4|}} \sin \frac{t(4-q^2)}{2\sqrt{|q^4-4|}}, \\ C(t) &= \cos \frac{t(4-q^2)}{2\sqrt{|q^2-4|}} - \frac{q}{\sqrt{|q^2-4|}} \sin \frac{t(4-q^2)}{2\sqrt{|q^4-4|}}. \end{aligned}$$

Соответствующее аффинное преобразование

$$\begin{aligned} (x^1, x^2, x^3) &\rightarrow (x^1 + t_1, x^2, x^3), \\ (x^1, x^2, x^3) &\rightarrow (x^1, x^2 + t_2, x^3), \\ (x^1, x^2, x^3) &\rightarrow (e^{\frac{-t_3 q}{2}} [A(t_3)x^1 + 2B(t_3)x^2], e^{\frac{-t_3 q}{2}} [-2B(t_3)x^1 + C(t_3)x^2], x^3 + t_3). \end{aligned} \quad (14)$$

Таким образом, группа Ли G_3VII имеет следующее аффинное представление:

$$\begin{cases} \bar{x}^1 = e^{\frac{-t_3 q}{2}} [A(t_3)x^1 + 2B(t_3)x^2] + t_1, \\ \bar{x}^2 = e^{\frac{-t_3 q}{2}} [-2B(t_3)x^1 + C(t_3)x^2] + t_2, \\ \bar{x}^3 = x^3 + t_3. \end{cases} \quad (15)$$

Теорема доказана. ■

ЛИТЕРАТУРА

1. Yamaguchi S. On complete affinely flat structures of some solvable Lie groups // *Memories of the Faculty of Science, Kyushu Univ., Ser. A.* 1979. V. 33, No. 2. P. 209–218.
2. Yamaguchi S. Supplements to "On complete affinely flat structures of some solvable Lie groups" // *Memories of the Faculty of Science, Kyushu Univ., Ser. A.* 1979. V. 33, No. 2. P. 219–223.
3. Петров А.З. Пространства Эйнштейна. М.: ФМ, 1961.

**AFFINE REPRESENTATIONS AS LIE GROUPS OF TRANSFORMATIONS
OF THREE-DIMENSIONAL SOLVABLE LIE GROUPS**

A.K. Guts

Dr.Sc. (Phys.-Math.), Professor, e-mail: guts@omsu.ru

G.B. Goldina

Student, e-mail: zoloto13@list.ru

A.N. Kabanov

Ph.D. (Phys.-Math.), Associate Professor, e-mail: kabanovan@omsu.ru

Dostoevsky Omsk State University, Omsk, Russia

Abstract. Using the Yamaguchi method, we give the representations as Lie groups of affine transformations of all three-dimensional solvable Lie groups.

Keywords: Affine structures, affine representation, three-dimensional solvable Lie groups.

Дата поступления в редакцию: 30.04.2018

УСТОЙЧИВОСТЬ НЕОДНОРОДНЫХ ТРЁХСЛОЙНЫХ СТЕРЖНЕЙ ПРИ НЕРАВНОМЕРНОМ ПОЛЕ ТЕМПЕРАТУРЫ В НЕЛИНЕЙНО УПРУГОЙ СРЕДЕ

Р.Э. Мамедли

доктор философии по математике, e-mail: prog-nv@mail.ru

Нижевартовский государственный университет, Нижневартовск, Россия

Аннотация. В статье исследуется задача устойчивости трёхслойных неоднородных прямолинейных стержней на нелинейно упругом основании под действием сжимающих нагрузок. Здесь предполагается, что стержень находится в неравномерном температурном поле и модули упругости материала слоев зависят от температуры. Для упругого основания принимается нелинейный модель и предполагается, что гипотеза плоских сечений справедлива для всей толщины элемента стержня. В общем виде получено уравнение устойчивости рассматриваемого стержня и для конкретного случая найдена формула для определения критической нагрузки.

Ключевые слова: неоднородный трёхслойный стержень, температура, устойчивость, критическая нагрузка.

Введение

Многослойные стержни часто используются в качестве несущих элементов во многих сложных конструкциях, работающих в различных режимах нагружения. Такие конструкции в некоторых случаях находятся на нелинейно упругом основании и в неравномерном температурном поле.

В работах [1–4] исследована устойчивость однослойных и многослойных стержней при нормальной температуре и под действием высокой температуры.

В данной работе исследуется задача устойчивости трёхслойных неоднородных стержней, которые находятся в неравномерном температурном поле и на нелинейно упругом основании.

Постановка задачи

Рассмотрим устойчивость трёхслойного неоднородного стержня на нелинейно упругом основании, имеющем две оси симметрии поперечного сечения, сжимаемого по концам силами P при неравномерном нагреве.

Координатная система выбрана следующим образом: оси OY и OZ находятся в поперечном сечении стержня; ось OX направлена по оси стержня.

Здесь предполагается, что температура в каждом слое является функцией координаты толщины (т. е. $T_i = T_i(z)$).

Слои стержня изготовлены из различных неоднородных материалов, и модули упругости зависят от координат и температуры следующим образом:

$$\begin{aligned} E_1(z) &= E_{10} - \beta_1(T - T_0), \\ E_2(z) &= E_{20} - \beta_2(T - T_0), \\ E(z) &= E_0 - \beta(T - T_0). \end{aligned} \quad (1)$$

В возмущённом состоянии стержня связь между приращениями напряжений и деформаций будет иметь вид:

$$\begin{aligned} \Delta\sigma &= E(z)\Delta\varepsilon, \quad \frac{-h_0}{2} \leq z \leq \frac{h_0}{2}, \\ \Delta\sigma^1 &= E_1(z)\Delta\varepsilon, \quad \frac{-h_0}{2} - h_1 \leq z \leq \frac{-h_0}{2}, \\ \Delta\sigma^2 &= E_2(z)\Delta\varepsilon, \quad \frac{h_0}{2} \leq z \leq \frac{h_0}{2} + h_1. \end{aligned} \quad (2)$$

Здесь h_1, h_2 — толщины соответствующих слоёв. Предположим, что гипотеза плоских сечений справедлива для всей толщины стержня т. е.

$$\Delta\varepsilon = e_0 + z\kappa. \quad (3)$$

Здесь e_0 — дополнительная деформация оси стержня, κ — кривизна центральной линии. В этом случае приращение усилий и момента определяются по формулам:

$$\Delta P = b \int_{\frac{-h_0}{2} - h_1}^{\frac{-h_0}{2}} \Delta\sigma^1 dz + b \int_{\frac{-h_0}{2}}^{\frac{h_0}{2}} \Delta\sigma dz + b \int_{\frac{h_0}{2}}^{\frac{h_0}{2} + h_2} \Delta\sigma^2 dz. \quad (4)$$

С учётом (1)–(3) из (4) находим

$$\begin{aligned} \Delta P &= [\overline{E_{10}}a_1^0 - \beta_1 A_1^0 + \overline{E_0}a^0 - \beta A^0 + \overline{E_{20}}a_2^0 - \beta_2 A_2^0]e + \\ &\quad + [\overline{E_{10}}a_1^1 - \beta_1 A_1^1 + \overline{E_0}a^1 - \beta A^1 + \overline{E_{20}}a_2^1 - \beta_2 A_2^1]\kappa, \\ \Delta M &= [\overline{E_{10}}a_1^1 - \beta_1 A_1^1 + \overline{E_0}a^1 - \beta A^1 + \overline{E_{20}}a_2^1 - \beta_2 A_2^1]e + \\ &\quad + [\overline{E_{10}}a_1^2 - \beta_1 A_1^2 + \overline{E_0}a^2 - \beta A^2 + \overline{E_{20}}a_2^2 - \beta_2 A_2^2]\kappa. \end{aligned} \quad (5)$$

В этих формулах введены следующие обозначения:

$$\begin{aligned} a_1^0 &= b \int_{\frac{-h_0}{2} - h_1}^{\frac{-h_0}{2}} dz, & a^0 &= b \int_{\frac{-h_0}{2}}^{\frac{h_0}{2}} dz, & a_2^0 &= b \int_{\frac{h_0}{2}}^{\frac{h_0}{2} + h_2} dz, \\ a_1^1 &= b \int_{\frac{-h_0}{2} - h_1}^{\frac{-h_0}{2}} z dz, & a^1 &= b \int_{\frac{-h_0}{2}}^{\frac{h_0}{2}} z dz, & a_2^1 &= b \int_{\frac{h_0}{2}}^{\frac{h_0}{2} + h_2} z dz, \end{aligned}$$

$$\begin{aligned}
a_1^2 &= b \int_{\frac{-h_0}{2}-h_1}^{\frac{-h_0}{2}} z^2 dz, & a^2 &= b \int_{\frac{-h_0}{2}}^{\frac{h_0}{2}} z^2 dz, & a_2^2 &= b \int_{\frac{h_0}{2}}^{\frac{h_0}{2}+h_2} z^2 dz, \\
A_1^0 &= b \int_{\frac{-h_0}{2}-h_1}^{\frac{-h_0}{2}} T(z) dz, & A_1^1 &= b \int_{\frac{-h_0}{2}-h_1}^{\frac{-h_0}{2}} T(z) z dz, & A_1^2 &= b \int_{\frac{-h_0}{2}-h_1}^{\frac{-h_0}{2}} T(z) z^2 dz, \\
A^0 &= b \int_{\frac{-h_0}{2}}^{\frac{h_0}{2}} T(z) dz, & A^1 &= b \int_{\frac{-h_0}{2}}^{\frac{h_0}{2}} T(z) z dz, & A^2 &= b \int_{\frac{-h_0}{2}}^{\frac{h_0}{2}} T(z) z^2 dz, \\
A_2^0 &= b \int_{\frac{h_0}{2}}^{\frac{h_0}{2}+h_2} T(z) dz, & A_2^1 &= b \int_{\frac{h_0}{2}}^{\frac{h_0}{2}+h_2} T(z) z dz, & A_2^2 &= b \int_{\frac{h_0}{2}}^{\frac{h_0}{2}+h_2} T(z) z^2 dz.
\end{aligned} \tag{6}$$

Получение уравнения устойчивости

Уравнения равновесия рассматриваемого стержня имеют вид

$$\Delta P = 0; \quad \Delta M = -PW \tag{7}$$

С учётом (5) из первого уравнения системы (7) получим

$$e = - \frac{[\overline{E_1} 0 a_1^1 - \beta_1 A_1^1 + \overline{E_0} a^1 - \beta A^1 + \overline{E_2} 0 a_2^1 - \beta_2 A_2^1]}{[\overline{E_1} 0 a_1^0 - \beta_1 A_1^0 + \overline{E_0} a^0 - \beta A^0 + \overline{E_2} 0 a_2^0 - \beta_2 A_2^0]} \varkappa. \tag{8}$$

С учётом (8) из (5) для приращения момента получим следующее выражение

$$\Delta M = \overline{K_2} I \varkappa. \tag{9}$$

Здесь $\overline{K_2} I$ — обобщённая характеристика твёрдости стержня определяется следующим образом

$$\begin{aligned}
\overline{K_2} I &= \left\{ [\overline{E_1} 0 a_1^2 - \beta_1 A_1^2 + \overline{E_0} a^2 - \beta A^2 + \overline{E_2} 0 a_2^2 - \beta_2 A_2^2] - \right. \\
&\quad \left. - \frac{[\overline{E_1} 0 a_1^1 - \beta_1 A_1^1 + \overline{E_0} a^1 - \beta A^1 + \overline{E_2} 0 a_2^1 - \beta_2 A_2^1]^2}{\overline{E_1} 0 a_1^0 - \beta_1 A_1^0 + \overline{E_0} a^0 - \beta A^0 + \overline{E_2} 0 a_2^0 - \beta_2 A_2^0} \right\}. \tag{10}
\end{aligned}$$

Учитывая, что $\varkappa = \frac{d^2 u}{dx^2}$, из второго уравнения системы (7) с учётом (9) получим

$$\frac{d^2 W}{dx^2} + R_2^2 W = 0; \quad R_2^2 = \frac{P}{\overline{K_2} I}. \tag{11}$$

Таким образом, уравнение устойчивости рассматриваемого стержня получено в виде (11).

Решение задачи

Для решения конкретных задач необходимо задавать вид зависимости модулей упругости от координат. Допустим, температура является линейной функцией координаты толщины, т. е.

$$T(z) = T_1 + \frac{T_2 z}{h_0} \quad (12)$$

Подставляя (12) в (6), получим:

$$\begin{aligned} a_1^0 &= bh_1, & a^0 &= bh_0, & a_2^0 &= bh_2 \\ a_1^1 &= -\frac{bh_0^2}{2}[\delta_1 + \delta_1^2], & a^1 &= 0, & \delta_1 &= \frac{h_1}{h_0}, \\ a_2^1 &= \frac{bh_0^2}{2}[\delta_2 + \delta_2^2], & \delta_2 &= \frac{h_2}{h_0}, \\ a_1^2 &= \frac{bh_0^3}{3} \left[-\frac{1}{8} + \left(\frac{1}{2} + \delta_1 \right)^3 \right], & a^2 &= \frac{bh_0}{12}, \\ a_2^2 &= \frac{bh_0^3}{3} \left[\left(\frac{1}{2} + \delta_2 \right)^3 - \frac{1}{8} \right], \\ A_1^0 &= bh_0 \left\{ T_1 \delta_1 + \frac{T_2}{2} \left[\frac{1}{4} - \left(\frac{1}{2} + \delta_1 \right)^2 \right] \right\}, \\ A_1^1 &= bh_0^2 \left\{ \frac{T_1}{2} \left[\frac{1}{4} - \left(\frac{1}{2} + \delta_1 \right)^2 \right] + \frac{T_2}{3} \left[-\frac{1}{8} + \left(\frac{1}{2} + \delta_1 \right)^3 \right] \right\}, \\ A_1^2 &= bh_0^3 \left\{ \frac{T_1}{3} \left[-\frac{1}{8} + \left(\frac{1}{2} + \delta_1 \right)^3 \right] + \frac{T_2}{4} \left[\frac{1}{16} - \left(\frac{1}{2} + \delta_1 \right)^4 \right] \right\}, \\ A^0 &= bT_1 h_0, & A^1 &= \frac{bT_2 h_0^2}{12}, & A^2 &= \frac{bT_1 h_0^3}{12}, \\ A_2^0 &= bh_0 \left\{ T_1 \delta_2 + \frac{T_2}{2} \left[\left(\frac{1}{2} + \delta_2 \right)^2 - \frac{1}{4} \right] \right\}, \\ A_2^1 &= bh_0^2 \left\{ \frac{T_1}{2} \left[\left(\frac{1}{2} + \delta_2 \right)^2 - \frac{1}{4} \right] + \frac{T_2}{3} \left[\left(\frac{1}{2} + \delta_2 \right)^3 - \frac{1}{8} \right] \right\}, \\ A_2^2 &= bh_0^3 \left\{ T_1/3 \left[\left(\frac{1}{2} + \delta_2 \right)^3 - \frac{1}{8} \right] + \frac{T_2}{4} \left[\left(\frac{1}{2} + \delta_2 \right)^4 - \frac{1}{16} \right] \right\} \end{aligned} \quad (13)$$

Подставляя выражения (13) в (10), определяем обобщённые характеристики жёсткости стержня и вычисляем критическую нагрузку на основе формулы (10).

Приведены вычисления при разных значениях параметров, и полученные результаты показаны на рисунке 1.

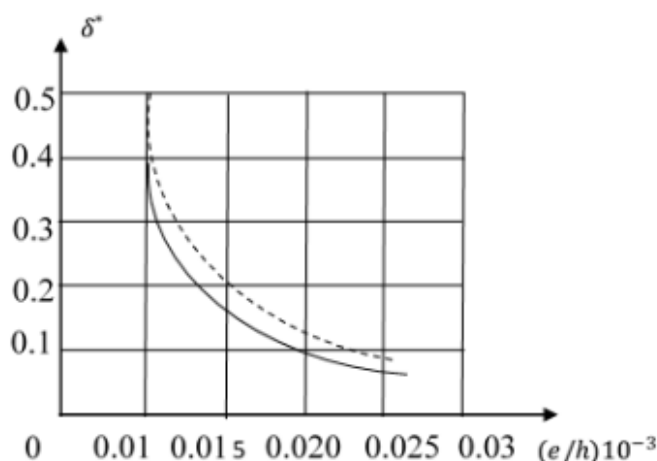


Рис. 1. Результаты

$$E_{10} = E_{20} = 2 \frac{10^6 \text{ кгс}}{\text{см}^2}; \quad E_0 = 1.8 \frac{10^6 \text{ кгс}}{\text{см}^2}; \quad \delta_1 = \delta_2 = 0.2; \quad \frac{h}{b} = 0.5;$$

$$\beta_1 = \beta_2 = 1000 \text{ кгс/см}^2 \text{град}; \quad \beta = 800 \text{ кгс/см}^2 \text{град};$$

$$T_0 = 300\text{K}; \quad T_1^1 = 300\text{K}; \quad T_1^2 = 500\text{K};$$

$$\delta^* = P/E_{10}bh_1.$$

ЛИТЕРАТУРА

1. Вольмир А.С. Устойчивость деформируемых систем. М.: Наука, 1967. 984 с.
2. Ломакин В.А. Теория упругости неоднородных тел. М.: Изд-во МГУ, 1976. 320 с.
3. Шаповалов Л.А. Влияние неравномерного нагрева на устойчивость сжатого стержня // ПММ. 1957. Т. XXII, № 1. С. 119–123.
4. Зубчанинов В.Г. Об упругопластической устойчивости слоистых стержней // Прикладная механика. 1970. Т. 6, № 2. С. 127–129.
5. Yang Y.B., Lin T.J., Len L.I. Thermal effect on the Postbuckling Behavior of an elastic or elasto-plastic truss // Journal of Mechanics. 2008. Т. 134, № 4. С. 330–338
6. Heydarpour A., Bradford M.A. Nonlinear Analysis of Composite Beams with Partial Interaction in steel Frame Structures at Elevated Temperature // Journal of Structural Engineering. 2010. Т. 136. С. 968–978.
7. Voshoughi A.R., Malekzadeh P., Banan Mo.R. Thermal postbuckling of laminated composite skew plates with temperature-dependent properties // Thin Walled structures. 2011. Т. 47, № 7. С. 804–811.
8. Vo T.P., Thai H.-T. Vibration and buckling of composite beams using refined shear deformation theory // International Journal of Mechanical Sciences. 2012. Т. 62, № 1. С. 67–76.

**STABILITY OF INHOMOGENEOUS THREE-LAYERED RODS
AT NON-UNIFORM TEMPERATURE FIELD IN A NONLINEARLY
ELASTIC MEDIUM**

R.E. Mamedli

Ph.D. (Math.), e-mail: prog-nv@mail.ru

Nizhnevartovsk State University, Nizhnevartovsk, Russia

Abstract. We study the problem of three-layered nonhomogeneous rectilinear rods on a nonlinear elastic foundation under the pressure of compressive loads in this article. It is assumed that the rod is in the uneven temperature field and the elasticity modules of the material layers depend on temperature. For the elastic foundation of nonlinear model is accepted and it is assumed that the hypothesis of plane sections is valid for the entire thickness of the element of the rod. In general, we achieve the steadiness equation of the considered rod, and a formula is found for determining the critical load in the certain case.

Keywords: nonhomogeneous three-layered rod, temperature, stability, critical load.

Дата поступления в редакцию: 01.03.2018

ПОСТРОЕНИЕ ДИСПЕРСИОННЫХ КОМПЛЕКСОВ ДЛЯ ОЦЕНКИ ЭФФЕКТИВНОСТИ ИММУНОТЕРАПИИ АЛЛЕРГИЧЕСКОЙ БРОНХИАЛЬНОЙ АСТМЫ

В.В. Гольяпин¹

к.ф.-м.н, доцент, e-mail: goltyapin@mail.ru

В.А. Шовин¹

научный сотрудник, e-mail: v.shovin@mail.ru

Е.В. Надей²

ассистент, e-mail: elenanadei@gmail.com

В.И. Совалкин²

д.м.н., профессор, e-mail: vsovalkin@gmail.com

Г.В. Нечаева²

д.м.н., профессор, e-mail: elenanadei@gmail.com

¹Институт математики им. С.Л. Соболева СО РАН (Омский филиал), Россия

²ФГБОУ ВО «Омский государственный медицинский университет»

Минздрава России

Аннотация. В данной статье проведена оценка эффективности аллерген-специфической иммунотерапии (АСИТ) аллергической бронхиальной астмы (БА), коморбидной с аллергическим ринитом (АР) и атопическим дерматитом (АтД) за трёхлетний период посредством однофакторного дисперсионного анализа многоградиционных признаков.

Ключевые слова: корреляционное отношение, организованные факторы, дискретные величины, дисперсионный анализ, бронхиальная астма, аллергия, аллергенспецифическая иммунотерапия.

Введение

Лечение и профилактика бронхиальной астмы (БА) является одним из актуальных приоритетов Всемирной организации здравоохранения в области неинфекционных заболеваний. С 1995 г. неоднократно проводился пересмотр рекомендаций ВОЗ по лечению больных БА, касающихся тактики противовоспалительной терапии.

В структуре бронхиальной астмы лёгкая форма болезни, которая успешно контролируется, составляет от 50 до 75 % пациентов. Больные с наличием среднетяжёлой и тяжёлой формы БА, коморбидной с аллергопатологией, составляют 25–40 % и вызывают большие трудности в достижении контроля над заболеванием.

К распространённым сопутствующим заболеваниям при бронхиальной астме относят, прежде всего, аллергический ринит (АР), который имеет ряд схожих

признаков с астмой, и атопический дерматит (АД). Медицинским сообществом активно обсуждается тема сочетания аллергического ринита и бронхиальной астмы. До 80-90 % больных БА страдают АР. В отсутствие терапии АР у больных БА возрастают частота её обострений и количество обращений за неотложной медицинской помощью, то есть не леченный АР негативно отражается на течении БА [1].

Аллергенспецифическая иммунотерапия (АСИТ), применяемая для лечения больных, страдающих атопическими IgE-опосредованными заболеваниями, с 1911 г. и в настоящее время остаётся одним из основных патогенетически обоснованных методов. Однако, несмотря на клинические исследования и мета-анализы, доказывающие эффективность АСИТ, она всё ещё применяется недостаточно и используется менее чем у 10 % пациентов с АР и БА. Оперативная группа ЕААСИ продолжает исследование стандартизованного метода эффективности АСИТ, который одновременно учитывает и симптомы заболевания, и потребность в противоаллергической терапии [2].

Многолетний положительный опыт, накопленный отечественными и зарубежными аллергологами по АСИТ атопических заболеваний, получил строго научное подтверждение её эффективности [3-5]. Согласно рекомендациям по АР и его влиянию на БА (ARIA), АСИТ используется для лечения среднетяжёлых и тяжёлых симптомов АР [6]. При этом можно достичь существенных клинических результатов и снизить потребность в симптоматических лекарственных препаратах, а также добиться сохранения клинической эффективности в течение нескольких лет после прекращения АСИТ [7].

АСИТ позволяет уменьшить симптомы и увеличить эффект противовоспалительной терапии БА, причём стероидсберегающий эффект АСИТ имеет первостепенное значение [8,9].

Всё ещё противоречивыми остаются мнения о потенциальной роли АСИТ в качестве терапевтического средства у пациентов с АтД. Однако имеются доказательства эффективности АСИТ у пациентов с выраженными симптомами АтД (индекс SCORAD > 50 баллов) [10, 11].

Таким образом, по литературным данным имеются указания на клиническую эффективность АСИТ атопической БА (БА), БА в сочетании с аллергическим ринитом (АР), изолированного атопического дерматита (АтД) [12–15]. Однако данные по клинической эффективности пациенториентированной АСИТ больных с БА средней степени тяжести, коморбидной с АР и АтД, при наличии поливалентной сенсibilизации отсутствуют.

В связи с этим была поставлена цель: оценить эффективность аллергенспецифической иммунотерапии в комбинации с базисной противовоспалительной терапией аллергической бронхиальной астмы средней тяжести, коморбидной с аллергическим ринитом и атопическим дерматитом при поливалентной сенсibilизации посредством дисперсионного анализа.

Математический аппарат дисперсионного анализа и алгоритм построения дисперсионных комплексов

Факторы — это любое воздействие или состояния, вариация которых может так или иначе влиять на вариации результативного признака.

Организация факторов заключается в подборе дискретных параметров или характеристик, которые оказывают статистическое влияние на результативный признак, при этом каждому изучаемому фактору придаётся несколько значений. В соответствии с этими значениями каждый фактор разбивается на несколько градаций, для каждой градации подбирается по принципу случайной выборки несколько объектов, у которых впоследствии и измеряется величина результативного признака [16, 17].

Для того, чтобы выяснить степень и достоверность влияния изучаемых, изучают и оценивают ту часть общей вариации, которая вызывается этими факторами. Основным коэффициентом, показывающим степень влияния того или иного организованного фактора, является квадрат корреляционного отношения [18]. Корреляционное отношение для непрерывных случайных величин рассчитывается по формуле (1), а для дискретных — по формуле (2).

$$\eta_{\xi\varphi}^2 = \frac{1}{\sigma_{\xi}^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (M(\xi|\varphi=x) - M(\xi))^2 p_{\xi\varphi} dy dx, \quad (1)$$

$$M(\xi|\varphi=x) = \int_{-\infty}^{\infty} y \frac{p_{\xi\varphi}(y,x)}{p_{\varphi}(x)} dy, \quad M\xi = \int_{-\infty}^{\infty} y p_{\xi}(y) dy,$$

$$\sigma_{\xi}^2 = M\xi^2 - (M\xi)^2 = \int_{-\infty}^{\infty} y^2 p_{\xi}(y) dy - \left(\int_{-\infty}^{\infty} y p_{\xi}(y) dy \right)^2,$$

$$\eta_{\xi\varphi}^2 = \frac{1}{\sigma_{\xi}^2} \sum_{i=0}^{n_{\varphi}} \sum_{j=0}^{n_{\xi}} (M(\xi|x_i) - M(\xi))^2 p(y_j, x_i), \quad (2)$$

$$M(\xi|x_i) = \sum_{j=0}^{n_{\xi}} y_j p(y_j|x_i), \quad M\xi = \sum_{j=0}^{n_{\xi}} y_j p(y_j),$$

$$\sigma_{\xi}^2 = M\xi^2 - (M\xi)^2 = \sum_{j=0}^{n_{\xi}} y_j^2 p(y_j) - \left(\sum_{j=0}^{n_{\xi}} y_j p(y_j) \right)^2.$$

Дисперсионный комплекс — это совокупность градаций с привлечёнными для исследования дискретными переменными или характеристиками, средними по каждой градации (частные средние) и по всему комплексу (общая средняя). Если изучается действие одного фактора, комплекс называется однофакторным, двух факторов — двухфакторным, трёх и более факторов — многофакторным.

Основными величинами однофакторного дисперсионного комплекса являются:

C_ξ — общая вариационная мера результативного признака;

C_x — факториальная мера вариации результативного признака;

C_z — случайная вариационная мера результативного признака.

В целях компактной записи основных формул дисперсионного комплекса введём следующие вспомогательные величины:

$$S_1 = \sum_{j=0}^g \sum_{i=0}^k f_{ij} \xi_i, \quad (3)$$

$$S_2 = \sum_{j=0}^g \sum_{i=0}^k f_{ij} \xi_i^2, \quad (4)$$

$$S_3 = \sum_{j=0}^g \frac{\left(\sum_{i=0}^k f_{ij} \xi_i \right)^2}{n_j}, \quad (5)$$

где ξ_i — значение результативного признака (дискретная случайная величина), изменяющегося в результате статистического влияния организованных и неорганизованных факторов, $i = 0, \dots, k$, k — количество возможных исходов результативного признака, а f_{ij} — значение частоты результативного признака ξ_i при j -ой градации, g — количество всевозможных градаций организованного факта. Как будет видно далее, в рамках построения дисперсионного комплекса квадрат корреляционного отношения сводится к отношению соответствующих вариационных мер.

Используя выше введённые обозначения, запишем для вариационных мер основные формулы однофакторного дисперсионного комплекса:

$$C_\xi = C_z + C_x;$$

$$C_\xi = S_2 - H; \quad (6)$$

$$C_z = S_2 - S_3; \quad (7)$$

$$C_x = S_3 - H. \quad (8)$$

Если любую вариационную меру разделить на общую вариационную меру, то получим коэффициент влияния фактора или сочетания факторов, равный квадрату корреляционного отношения:

$$\eta_z^2 = \frac{C_z}{C_\xi}, \quad \eta_x^2 = \frac{C_x}{C_\xi}. \quad (9)$$

А если поделить вариационную меру на соответствующее число степеней свободы, получим дисперсию:

$$\sigma_z^2 = \frac{C_z}{n - g}, \quad \sigma_x^2 = \frac{C_x}{g - 1}, \quad \sigma_\xi^2 = \frac{C_\xi}{n - 1}. \quad (10)$$

Для проверки достоверности влияния фактора или сочетания факторов используется критерий Фишера в виде отношения соответствующей факториальной дисперсии к случайной дисперсии:

$$F_x = \frac{\sigma_x^2}{\sigma_z^2}. \quad (11)$$

Алгоритм построения однофакторного дисперсионного комплекса.

1. Из исходной таблицы данных объект-свойство, размерности $n \times 2$, вычисляем значения f_{ij} .
2. По формулам (3)–(5) вычисляем вспомогательные величины S_1, S_2, S_3 .
3. Используя вычисленные S_1, S_2, S_3 по формулам (6)–(8), вычисляем вариационные меры C_x, C_y, C_z .
4. Находим квадрат корреляционного отношения η_x^2 по формуле (9), определяющий степень влияния организованного фактора.
5. Вычисляем общую σ_ξ^2 , факториальную σ_x^2 и случайную σ_z^2 дисперсии по формуле (10).
6. Для проверки достоверности влияния проводим расчёт критерия Фишера F_x , как отношение факториальной дисперсии к общей дисперсии (формула 11).
7. Проводим расчёт уровня значимости α по F_x при степенях свободы $g - 1$ и $n - g$.

Вычислительный эксперимент и полученные результаты

Тяжесть течения БА, АР и АтД на момент обследования представляли в баллах, субъективно оцениваемых самими больными. Учитывали выраженность приступов удушья, чихания, заложенности носа, зуда кожи и интенсивность кожных проявлений по шкале: 0 – отсутствие клинических симптомов, 1 – лёгкая степень выраженности симптомов, 2 – средняя (умеренная) степень выраженности, 3 – выраженные проявления. В исследовании участвовали 119 пациентов в возрасте от 18 до 32 лет. Результаты вычислительного эксперимента представлены ниже следующими таблицами.

Однофакторные дисперсионные комплексы рассчитывались и анализировались по каждому симптому в отдельности, за организованный фактор была выбрана АСИТ, проводимая в течении трёх лет в одной из групп и не проводимая в другой, получающей базисную противовоспалительную терапию без АСИТ.

Таблица 1. Таблица распределения симптома (η) «ринорея» по годам (ξ) у пациентов коморбидной аллергической БА

Распределение симптома «ринорея» по годам									
1-ая группа					2-ая группа				
$\xi \setminus \eta$	0	1	2	3	$\xi \setminus \eta$	0	1	2	3
1	2	28	25	33	1	0	10	9	12
2	45	35	8	0	2	9	11	8	3
3	68	19	1	0	3	17	12	2	0
4	77	11	0	0	4	19	6	6	0

Таблица 2. Таблица двух однофакторных дисперсионных комплексов, рассчитанных по симптому «ринорея» у пациентов аллергической БА, коморбидной с АР и АтД

Аллергический симптом – ринорея							
1-ая группа				2-ая группа			
S_1	S_2	S_3	H	S_1	S_2	S_3	H
260	526	391.9545	192.0454	134	274	192.6451	144.8064
C_ξ	C_x	C_z	η_x	C_ξ	C_x	C_z	η_x
333.9545	199.9090	134.0454	0.7737	129.1935	47.8387	81.3548	0.6085
σ_ξ^2	σ_x^2	σ_z^2	η_x^2	σ_ξ^2	σ_x^2	σ_z^2	η_x^2
0.9514	66.6363	0.3851	0.5986	1.2601	15.9462	0.67795	0.3702
F	α	$g - 1$	$n - g$	F	α	$g - 1$	$n - g$
172.9969	1	3	349	23.5210	1	3	90

Таблица 3. Таблица распределения симптома (η) «заложенность носа» по годам (ξ) у пациентов с аллергической БА, коморбидной с АР и АтД

Распределение симптома «заложенность носа» по годам лечения									
1-ая группа					2-ая группа				
$\xi \setminus \eta$	0	1	2	3	$\xi \setminus \eta$	0	1	2	3
1	0	0	35	53	1	0	0	11	20
2	14	48	20	6	2	2	17	12	0
3	42	37	9	0	3	3	18	10	0
4	57	25	6	0	4	7	19	5	0

Таблица 4. Таблица двух однофакторных дисперсионных комплексов, рассчитанных по симптому «заложенность носа» у пациентов аллергической БА, коморбидной с АР и АтД

Аллергический симптом – «заложенность носа»							
1-ая группа				2-ая группа			
S_1	S_2	S_3	H	S_1	S_2	S_3	H
427	921	773.5340	517.9801	190	386	344.8387	291.1290
C_ξ	C_x	C_z	η_x	C_ξ	C_x	C_z	η_x
403.01988	255.5539	147.4659	0.7963	94.8709	53.7096	41.1612	0.7524
σ_ξ^2	σ_x^2	σ_z^2	η_x^2	σ_ξ^2	σ_x^2	σ_z^2	η_x^2
1.1482	85.1846	0.4237	0.63409	0.7713	17.9032	0.343011	0.5661
F	α	$g - 1$	$n - g$	F	α	$g - 1$	$n - g$
201.0245	1	3	349	52.1943	1	3	90

Таблица 5. Таблица распределения симптома (η) «приступ удушья» по годам (ξ) у пациентов с аллергической БА, коморбидной с АР и АтД

Распределение симптома «приступ удушья» по годам лечения									
1-ая группа					2-ая группа				
$\xi \setminus \eta$	0	1	2	3	$\xi \setminus \eta$	0	1	2	3
1	0	0	32	56	1	0	0	9	22
2	18	46	20	4	2	1	10	14	6
3	38	41	9	0	3	1	12	17	1
4	61	23	4	0	4	1	12	11	7

Таблица 6. Таблица двух однофакторных дисперсионных комплексов, рассчитанных по симптому «приступ удушья» у пациентов с аллергической БА, коморбидной с АР и АтД

Аллергический симптом – «приступ удушья»							
1-ая группа				2-ая группа			
S_1	S_2	S_3	H	S_1	S_2	S_3	H
420	910	771.25	501.1363	244	562	503.8064	480.1290
C_ξ	C_x	C_z	η_x	C_ξ	C_x	C_z	η_x
408.8636	270.1136	138.75	0,8128	81.8709	23.6774	58.1935	0,53777
σ_ξ^2	σ_x^2	σ_z^2	η_x^2	σ_ξ^2	σ_x^2	σ_z^2	η_x^2
1.1648	90.0378	0.3987	0.6606	0.6656	7.8924	0.343011	0.4849
F	α	$g - 1$	$n - g$	F	α	$g - 1$	$n - g$
225.8247	1	3	349	16.2749	1	3	90

Таблица 7. Таблица распределения симптома (η) «зуд кожи» по годам (ξ) у пациентов с аллергической БА, коморбидной с АР и АтД

Распределение симптома «зуд кожи» по годам лечения									
1-ая группа					2-ая группа				
$\xi \setminus \eta$	0	1	2	3	$\xi \setminus \eta$	0	1	2	3
1	0	22	54	12	1	0	9	15	7
2	36	39	12	1	2	0	15	14	2
3	61	25	2	0	3	6	9	13	3
4	67	21	0	0	4	3	11	12	5

Таблица 8. Таблица двух однофакторных дисперсионных комплексов, рассчитанных по симптому «зуд кожи» у пациентов с аллергической БА, коморбидной с АР и АтД

Аллергический симптом – «зуд кожи»							
1-ая группа				2-ая группа			
S_1	S_2	S_3	H	S_1	S_2	S_3	H
282	496	377.2045	225.9204	203	413	336.6774	332.3306
C_ξ	C_x	C_z	η_x	C_ξ	C_x	C_z	η_x
270.0795	151.2840	118.7954	0,7484	80.6693	4.3467	76.3225	0,2321
σ_ξ^2	σ_x^2	σ_z^2	η_x^2	σ_ξ^2	σ_x^2	σ_z^2	η_x^2
0.7694	50.4280	0.3413	0.56014	0.6558	1.4489	0.6360	0.053884
F	α	$g - 1$	$n - g$	F	α	$g - 1$	$n - g$
225.8247	1	3	349	2.2781	1	3	90

Таблица 9. Таблица распределения симптома (η) «выраженность кожных проявлений» по годам (ξ) у пациентов с аллергической БА, коморбидной с АР и АтД

Распределение симптома «выраженность кожных проявлений» по годам лечения									
1-ая группа					2-ая группа				
$\xi \setminus \eta$	0	1	2	3	$\xi \setminus \eta$	0	1	2	3
1	0	4	70	14	1	0	0	30	1
2	17	46	17	8	2	2	10	16	3
3	45	41	2	0	3	1	12	16	2
4	60	27	1	0	4	1	10	17	3

Таблица 10. Таблица двух однофакторных дисперсионных комплексов, рассчитанных по симптому «выраженность кожных проявлений» у пациентов с аллергической БА, коморбидной с АР и АтД

Аллергический симптом – «выраженность кожных проявлений»							
1-ая группа				2-ая группа			
S_1	S_2	S_3	H	S_1	S_2	S_3	H
364	676	548.6136	376.4090	217	429	383.1935	379.75
C_ξ	C_x	C_z	η_x	C_ξ	C_x	C_z	η_x
299.5909	172.2045	127.3863	0,7581	49.25	3.4435	45.8064	0,2644
σ_ξ^2	σ_x^2	σ_z^2	η_x^2	σ_ξ^2	σ_x^2	σ_z^2	η_x^2
0.8535	57.4015	0.3660	0.5748	0.4004	1.1478	0.3817	0.0699
F	α	$g - 1$	$n - g$	F	α	$g - 1$	$n - g$
156.8121	1	3	349	3.007	1	3	90

Таблица 11. Таблица распределения уровней фармакотерапии (η) по годам (ξ) у пациентов с аллергической БА, коморбидной с АР и АтД

Распределение уровней фармакотерапии по годам лечения							
1-ая группа				2-ая группа			
$\xi \setminus \eta$	0	1	2	3	$\xi \setminus \eta$	2	3
1	0	0	77	11	1	24	7
2	2	21	65	0	2	29	2
3	33	15	40	0	3	29	2
4	53	3	32	0	4	23	8

Таблица 12. Таблица двух однофакторных дисперсионных комплексов, рассчитанных по уровням к требованию в фармакотерапии у пациентов с аллергической БА, коморбидной с АР и АтД

Уровни требуемой фармакотерапии							
1-ая группа				2-ая группа			
S_1	S_2	S_3	H	S_1	S_2	S_3	H
500	994	810.0454	3710.2272	267	591	575.9032	574.9112
C_ξ	C_x	C_z	η_x	C_ξ	C_x	C_z	η_x
283.7727	99.8181	183.9545	0,5930	16.0887	0.9919	15.0967	0,2483
σ_ξ^2	σ_x^2	σ_z^2	η_x^2	σ_ξ^2	σ_x^2	σ_z^2	η_x^2
0.8084	33.2727	0.5286	0.351754	0.1308	0.9919	0.123744	0.0616
F	α	$g - 1$	$n - g$	F	α	$g - 1$	$n - g$
62.9444	1	3	349	3.007	0.994575	2	90

Анализ полученных результатов и выводы

В 1-ой группе (пациенты на фоне базисной противовоспалительной терапии (БПВТ) получали аллергенспецифическую иммунотерапию (АСИТ)) по всем клиническим проявлениям бронхиальной астмы, аллергического ринита и атопического дерматита к окончанию 3-хлетнего периода наблюдения прослеживаются достоверные различия по отношению к исходным показателям.

При этом во 2-ой группе отмечено отсутствие статистически значимых различий частоты обострений БА, степени выраженности кожного зуда, объёма фармакотерапии, показателей ОФВ₁ и АСТ к окончанию третьего года терапии. В частности: в динамике ежегодно в 2 раза увеличивается количество пациентов с купированием ринореи как одного из симптомов АР, который к окончанию третьего года комбинированной терапии (БПВТ+АСИТ) полностью исчезает у 87,5 % пациентов, а у 12 % остаётся лёгкая эпизодическая его форма.

У пациентов 2-ой группы также отмечен хороший эффект на фоне продолжающейся БПВТ, к концу третьего года базисной противовоспалительной терапии ринорея купирована в 61 % случаях, лёгкая её форма зарегистрирована у 19 % больных, однако ещё у 19 % пациентов остаётся средне-выраженная её интенсивность. Статистически значимое снижение выраженности симптомов заложенности носа у пациентов 1-ой группы по сравнению с пациентами 2-ой группы проявлялось к концу второго курса терапии и сохранялось к окончанию 3-го курса. У половины пациентов обеих групп к концу второго года лечения достигнуто купирование заложенности носа. К окончанию исследования процент отличных результатов в 1-ой группе увеличился с 50,4 % до 64,8 %, у 28 % сохранялись лёгкие проявления, в 7 % случаев — заложенность носа была средней степени выраженности, требующая непрерывной противовоспалительной терапии. При этом у 35,8 % пациентов 2-ой группы к окончанию исследования заложенность носа возобновилась, лёгкая её форма регистрировалась в 61 % случаев, среднетяжёлая — в 16 %.

Динамика частоты приступов бронхиальной астмы показательна спустя два года терапии у пациентов обеих групп. У 52 % больных 1-ой группы и 33 % — 2-ой группы достигнуто купирование приступов удушья. В последующем количество пациентов с полным купированием удушья увеличивается в 1-ой группе и к окончанию 3-го года терапии достигает 69,4 %, при этом во 2-ой группе число пациентов с отсутствием приступов бронхиальной астмы резко уменьшается и достигает 3 %, практически равномерно распределяясь в лёгкую, умеренную и выраженную степень (38,7 %, 35,5 %, 22,6 % соответственно).

Данные по анализу интенсивности зуда кожи свидетельствуют об эффективности проводимой терапии у пациентов 1-ой группы — так, к концу 3 курса терапии 76,1 % пациентов отметили полное его купирование, при этом 23,9 % человек беспокоило эпизодическое его возобновление. У пациентов 2-ой группы добиться купирования зуда удалось лишь в 10 % случаях, в 35,5 % — зуд носил эпизодический характер, однако у 38,7 и 16 % больных кожный зуд сохранялся с умеренной и высокой степенью выраженности.

Использование однофакторного дисперсионного анализа позволило зафиксировать положительную динамику уменьшения клинических проявлений БА и коморбидной с ней патологии (АР и АтД) на протяжении всего трёхлетнего периода АСИТ в комплексе с БПВТ. Особый интерес вызвала возможность проследить изменения качественных признаков по степени выраженности, а не только по наличию или отсутствию симптома.

Комбинированная терапия (БПВТ+ АСИТ) у пациентов с аллергической БА средней степени тяжести, коморбидной с АР и АтД, при наличии поливалентной сенсибилизации к бытовым и пыльцевым аллергенам, через 3 года от начала комплексной терапии позволяет достичь полного купирования клинических проявлений БА (86,3 %), с отменой БПВТ в 63,6 % случаев.

Улучшение течения коморбидной патологии на фоне комбинированной БПВТ и АСИТ к окончанию 3-го курса терапии даёт максимальный клинический эффект в виде отсутствия обострений кожных проявлений (76,1 %) и полного купирования назальной обструкции (64,8 %).

ЛИТЕРАТУРА

1. Bousquet J., Gaugris S., Kocevar V.S. et.al. Increased risk of asthma attacks and emergency visits among asthma patients with allergic rhinitis: a subgroup analysis of the investigation of montelukast as a partner agent for complementary therapy // Clin Exp Allergy. 2005. V. 35(6). P. 723–727. DOI: 10.1111/j.1365-2222.2005.02251.x.
2. Pfaar O., Demoly P., Gerth van Wijk R. et al.; European Academy of Allergy and Clinical Immunology. Recommendations for the standardization of clinical outcomes used in allergen immunotherapy trials for allergic rhinoconjunctivitis: an EAACI Position Paper // Allergy. 2014. V. 69(7). P. 854–67. DOI: 10.1111/all.12383.
3. Павлова К.С., Курбачева О.М. Патогенетическая терапия аллергических заболеваний: возможности АСИТ в России // Эффективная фармакотерапия. Аллергология и иммунология. 2012. № 2(2). P. 10–35.
4. Козулина И.Е., Павлова К.С., Курбачева О.М. Клиническая эффективность подкожной и сублингвальной аллергенспецифической иммунотерапии аллергического ринита и конъюнктивита // Российский аллергологический журнал. 2016. № 6. С. 63–69.
5. Zielen S., Kardos P., Madonini E. Steroid-sparing effects with allergen-specific immunotherapy in children with asthma: a randomized controlled trial // J. Allergy Clin Immunol. 2010. V. 126(5). P. 942–951. DOI: 10.1016/j.jaci.2010.06.002.
6. Brozek J.L., Bousquet J., Baena-Cagnani C.E. et al. Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines: 2010 revision // J. Allergy Clin Immunol. 2010. V. 126(3). P. 466–542. DOI: 10.1016/j.jaci.2010.06.047.
7. Canonica G.W., Cox L., Pawankar R. et al. Sublingual immunotherapy: World Allergy Organization position paper 2013 update // World Allergy Organ J. 2014. V.7(1). DOI: 10.1186/1939-4551-7-6.
8. Гущин И.С., Курбачева О.М. Аллергия и аллергенспецифическая иммунотерапия. М. : «Фармарус Принт Медиа», 2010. 228 с.
9. Черняк Б.А., Сукманская Е.О. Эффективность специфической иммунотерапии в сочетании с ингаляционными кортикостероидами при атопической бронхиальной астме среднетяжёлого течения // Аллергология. 1999. № 1. С. 4–10.

10. Compalati E., Rogkakou A., Passalacqua G. et al. Evidences of efficacy of allergen immunotherapy in atopic dermatitis: an updated review // *Curr Opin Allergy Clin Immunol*. 2012. V. 12(4). P. 427–460. DOI: 10.1097/ACI.0b013e328354e540.
11. Novak N., Bieber T., Hoffmann M. Efficacy and safety of subcutaneous allergen-specific immunotherapy with depigmented polymerized mite extract in atopic dermatitis // *J. Allergy Clin Immunol*. 2012. V. 130(4). P. :925–956. DOI: 10.1016/j.jaci.2012.08.004
12. Nelson H.S. Subcutaneous injection immunotherapy for optimal effectiveness // *Immunol Allergy Clin North Am*. 2011. V. 31(2). P. 211–237. DOI: 10.1016/j.iac.2011.02.010.
13. Jutel M., Agache I., Bonini S. et al. International consensus on allergy immunotherapy // *J. Allergy Clin Immunol*. 2015. V. 136(3). P. 556–624. DOI: 10.1016/j.jaci.2015.04.047.
14. Abramson M.J., Puy R.M., Weiner J.M. Injection allergen immunotherapy for asthma // *Cochrane Database Syst Rev*. 2010. V. 4(8). DOI: 10.1002/14651858.
15. Jacobsen L., Wahn U., Bilo M.B. Mechanisms of allergen-specific immunotherapy // *Clin Transl Allergy*. 2012. V. 2. DOI: 10.1186/2045-7022-2-8.
16. Плохинский Н.А. Алгоритмы биометрии. М. : Изд-во Моск. гос. ун-та, 1980. 150 с.
17. Плохинский Н.А. Биометрия. Новосибирск : СО АН СССР, 1961. 364 с.
18. Крамер Г. Математические методы статистики. М. : Мир, 1975. 648 с.

BUILDING OF DISPERSION COMPLEXES FOR ESTIMATION OF EFFICIENCY OF IMMUNOTHERAPY OF ALLERGIC BRONCHIAL ASTHMA

V.V. Golyapin¹

Ph.D. (Phys.-Math.), Associate Professor, e-mail: golyapin@mail.ru

V.A. Shovin¹

Researcher, e-mail: v.shovin@mail.ru

E.V. Nadey²

Assistant, e-mail: elenanadei@gmail.com

V.I. Sovalkin²

M.D., Professor, e-mail: vsovalkin@mail.ru

G.I. Nechaeva²

M.D., Professor, e-mail: profnechaeva@yandex.ru

¹Omsk Branch of the Federal State budget institution S.L. Sobolev Institute of Mathematics of Siberian Branch of the RAS, Omsk, Russia

²Federal State Budget Educational Institution of Higher Education "Omsk State Medical University" of the Ministry of Health of the Russian Federation

Abstract. In this article, the efficacy of allergen-specific immunotherapy (ASIT) of allergic bronchial asthma (BA), comorbid with allergic rhinitis (AP) and atopic dermatitis (ATD) over a three-year period by means of a single-factor variance analysis of multigradation features was evaluated

Keywords: correlation relation, organized factors, discrete values, variance analysis, bronchial asthma, allergy, allergen-specific immunotherapy.

Дата поступления в редакцию: 01.03.2018

НЕЛИНЕЙНЫЕ СТРУКТУРНЫЕ УРАВНЕНИЯ И КВАДРАТИЧНЫЙ ФАКТОРНЫЙ АНАЛИЗ

В.А. Шовин

научный сотрудник, e-mail: v.shovin@mail.ru

Институт математики им. С.Л. Соболева СО РАН (Омский филиал), Омск, Россия

Аннотация. Роль статьи заключается в разработке нового алгоритма (метода) факторного анализа, включающего в себя продвинутую модель взаимоотношений факторов и исходных переменных, когда в модели учитываются не только линейные слагаемые, но и квадратичные, т.е. разложение до второй степени. Такая модель относится к классу нелинейных структурных моделей. Алгоритм расчёта позволяет строить уточняющую модель, исследовать влияние нелинейной составляющей в факторном влиянии. Также разработана соответствующая программа с графическим интерфейсом пользователя, способная проводить квадратичный факторный анализ новых данных по загрузке новых данных. Данная статья носит как прикладной, так и фундаментальный характер. Программа и соответствующий вычислительный алгоритм протестированы на реальных данных, что позволило сравнить предыдущие расчёты по этим данным для линейного факторного анализа с предлагаемым в статье квадратичным. На базе численных методов нелинейной оптимизации с условиями разработан способ нахождения оценки параметров нелинейных структурных уравнений. В качестве критерия оптимальных параметров структурных уравнений выбрана минимизация невязок структурных уравнений модели. Проведено тестирование данного метода для модели квадратичного факторного анализа артериальной гипертензии начальной стадии с соответствующей программной реализацией.

Ключевые слова: артериальная гипертензия, факторный анализ, структурные уравнения.

Введение

Структурные уравнения являются формой описания зависимостей между измеряемыми и латентными (не измеряемыми) переменными исследуемого объекта. Метод моделирования отношений между несколькими измеренными и латентными переменными оформился в 1970-х гг. в работах статистиков К. Йорескога и Д. Сёрбома [1], социологов Г. Блэлока, О. Дункана [2, 3], эконометристов А. Голдбергера [4] и психометриста П. Бентлера [5]. В общем случае такие зависимости могут иметь нелинейный характер функций модели.

Целью данной работы является поиск способа оценки параметров нелинейных структурных уравнений с дополнительными условиями на параметры структурной модели и построение структурной модели артериальной гипертензии.

Задачи данного исследования представлены следующими пунктами:

- представить задачу оценки параметров структурных уравнений как задачу оптимизации с условиями;
- программно реализовать методы оптимизации с условиями и адаптировать функцию критерия оптимизации от варьируемых параметров структурной модели;
- протестировать данный метод оценки параметров структурных уравнений на примере квадратичной факторной модели для данных артериальной гипертензии начальной стадии с известной факторной структурой;
- реализовать данный метод оценки факторных структур программно с графическим интерфейсом пользователя.

В качестве экспериментальных данных выступали различные параметры, характеризующие состояние пациентов с артериальной гипертензией. Поскольку экспериментальные данные представлены выборкой значений измеряемых переменных у различных исследуемых объектов, для оценки параметров и значений латентных переменных модели, задаваемой структурными уравнениями, может быть использован критерий минимальных невязок как сумма невязок модели вычисленных для всей выборки различных объектов. Дополнительно на параметры и значения латентных переменных могут быть заданы ограничительные условия. Для решения задачи минимизации невязок модели предлагается использовать методы нелинейной оптимизации с условиями: метод конфигураций (метод Хука-Дживса), метод случайного поиска (метод Монте-Карло). Ограничения, накладываемые на значения параметров и латентных переменных модели, учитываются с помощью метода штрафных функций. Тестирование метода оценки параметров структурных уравнений проводилось на модели квадратичного факторного анализа, как частного нелинейного случая системы структурных уравнений.

Квадратичный факторный анализ является обобщением линейного факторного анализа для учёта нелинейных зависимостей между исходными переменными. Учёт влияния нелинейной факторной составляющей позволяет ограничиться меньшим количеством факторов при той же самой ошибке модели. Иначе говоря, факторы являются нелинейными, чтобы их описать посредством линейных факторов, потребовалось их большее количество. Особенностью нелинейного фактора является его нелинейный характер, который может быть вызван специальным режимом, при котором он функционирует, а не влиянием комбинаций линейных факторов.

Учёт нелинейных зависимостей между исходными переменными в модели квадратичного факторного анализа происходит посредством ввода квадратичных слагаемых равных произведению различных пар факторов. Те слагаемые, которые имеют преимущественное влияние на объяснение вариаций значений исходных переменных, говорят об одновременном влиянии соответствующей

пары факторов. Одновременное воздействие на два таких фактора приводит к синергетическому влиянию на исходные показатели, т. е. влияние не равно сумме влияний отдельных факторов по отдельному их воздействию. Возможна и отрицательная синергия, когда два фактора, влияющих одновременно, приводят к меньшему воздействию на исходные переменные, чем сумма их отдельных влияний.

1. Факторный анализ как частный случай структурных уравнений

В теории структурных уравнений используются следующие типы матриц.

Матрица $Z \overset{m \times n}{\leftrightarrow} z_{ij}$ – матрица значений измеряемых переменных у исследуемых объектов или состояний объекта размерности $m \times n$, где m – число измеряемых параметров, n – число объектов или состояний объекта (объём выборки).

Матрица $P \overset{g \times n}{\leftrightarrow} p_{ij}$ – матрица значений латентных переменных объектов размерности $g \times n$, где g – число латентных параметров.

Матрица $A \overset{k \times s}{\leftrightarrow} a_{ij}$ – матрица параметров структурных уравнений размерности $k \times s$, где k – число структурных уравнений, s – число параметров в структурных уравнениях.

Система структурных уравнений задаётся в виде:

$$\begin{cases} f_1(a_{11}, a_{12}, \dots, a_{1s}; p_{1t}, p_{2t}, \dots, p_{gt}; z_{1t}, z_{2t}, \dots, z_{mt}) + \varepsilon_{1t} = 0, \\ f_2(a_{21}, a_{22}, \dots, a_{2s}; p_{1t}, p_{2t}, \dots, p_{gt}; z_{1t}, z_{2t}, \dots, z_{mt}) + \varepsilon_{2t} = 0, \\ \vdots \\ f_k(a_{k1}, a_{k2}, \dots, a_{ks}; p_{1t}, p_{2t}, \dots, p_{gt}; z_{1t}, z_{2t}, \dots, z_{mt}) + \varepsilon_{kt} = 0, \end{cases}$$

где $f_1, f_2, \dots, f_k$ – в общем случае нелинейные функции своих переменных, $\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{kt}$ – невязки модели для t -го объекта или состояния объекта.

На значения параметров и значения латентных переменных могут накладываться дополнительные условия в виде равенств и неравенств.

Оптимальными значениями параметров и латентных переменных считаются те значения, которые минимизируют абсолютные значения невязок модели и удовлетворяют всем дополнительным условиям.

В данной работе рассмотрен частный случай структурной модели – квадратичная факторная модель, описываемая следующими уравнениями [6, 7]:

$$\begin{cases} z_{1t} = a_{11}p_{1t} + a_{12}p_{2t} + \dots + a_{1g}p_{gt} + \sum_{i,j=1}^g b_{1ij}p_{it}p_{jt} + \varepsilon_{1t}, \\ z_{2t} = a_{21}p_{1t} + a_{22}p_{2t} + \dots + a_{2g}p_{gt} + \sum_{i,j=1}^g b_{2ij}p_{it}p_{jt} + \varepsilon_{2t}, \\ \vdots \\ z_{mt} = a_{m1}p_{1t} + a_{m2}p_{2t} + \dots + a_{mg}p_{gt} + \sum_{i,j=1}^g b_{mij}p_{it}p_{jt} + \varepsilon_{mt}, \end{cases} \quad (1)$$

где матрица $A \leftrightarrow a_{ij}$ называется матрицей факторной структуры размерности $m \times g$ весовых коэффициентов; m – число изучаемых параметров, g – число общих факторов; $B \leftrightarrow b_{kij}$ – матрица квадратичных коэффициентов.

На вид факторной структуры A и квадратичные коэффициенты налагаются дополнительные ограничения.

Общности переменных факторной структуры должны быть не больше 1, а также не меньше определённого порога значимости:

$$h_i = \sqrt{\sum_{k=1}^g a_{ik}^2} \leq 1. \quad (2)$$

Аналогично для квадратичных коэффициентов:

$$\sqrt{\sum_{ij=1}^g b_{kij}^2} \leq 1; \quad (3)$$

$$p_{ij} \leq 10. \quad (4)$$

Ограничение (2) является классическим ограничением линейного факторного анализа, когда норма вектора исходной переменной в факторном пространстве, т. н. общность, является ограниченной. Ограничения (2) и (3) на значения линейных и квадратичных коэффициентов факторной модели задаются в виде нестрогих равенств. Тем самым с факторной модели снимается сильная напряжённость и нагрузка, предоставляется большая гибкость и свобода связи параметров. Заданы лишь рамки факторной модели, чтобы факторная структура не «перекособочивалась» из-за нагрузки и напряжённости в другом месте. Ограничение (4) обосновано тем, что факторы p_i нормированы на их дисперсию D_i , и большая часть значений факторов находится за $3D_i$, в то время как значение фактора равное $10D_i$ – ещё более маловероятное событие.

Критерий оптимизации задаётся в следующем виде:

$$K = \sum_{t=1}^n \sum_{k=1}^m \varepsilon_{kt}^2 + \frac{1}{K_{int}(A)} + \frac{1}{K_{int}(B)}, \quad (5)$$

где K_{int} – критерий интерпретируемости, минимизация критерия K и учёт дополнительных условий на вид факторной структуры и квадратичных коэффициентов приводит к оптимальному решению для варьируемых значений элементов факторной структуры a_{ij} , квадратичных коэффициентов b_{kij} и факторов p_{ij} .

Получение интерпретируемого факторного решения связано с получением минимальной сложности исходных параметров, когда только одна факторная нагрузка переменной близка к 1, тогда как остальные близки к 0. Поэтому предлагается следующий критерий [8], непосредственно учитывающий это свойство:

$$K_{int}(V) = \sum_{i=1}^m \sum_{p=1}^g |v_i^{\max}| - |v_{ip}|, \quad (6)$$

где $v_i^{\max}$ – максимальная по модулю факторная нагрузка i -ой переменной факторной структуры.

Максимизация данного критерия приводит к тому, что максимальная факторная нагрузка переменной приближается к 1, тогда как остальные к 0.

Возможность взятия обратной величины критерия интерпретируемости в критерии оптимизации для превращения задачи максимизации в задачу минимизации обосновано использованием неградиентных методов оптимизации нулевого порядка.

Вычислительный алгоритм

Алгоритм построения квадратичной факторной модели.

1. Определение числа факторов числом $g < m$.
2. Определение начальных приближений матрицы A линейной части размерности $m \times g$, матрицы B квадратичной части размерности $m \times (g \cdot g)$ и матрицы P значений факторов размерности $g \times n$ случайными числами из диапазона $[-1; 1]$.
3. Минимизация критерия (4) суммы квадратов невязок структурных уравнений (1) и критерия интерпретируемости факторной структуры (5) как функций от независимых переменных матриц линейной A и квадратичной B частей факторной структуры и значений факторов P с ограничениями (2, 3) методом штрафных функций и методом конфигураций.

2. Методы оптимизации

Оптимизацию суммы квадратов невязок структурных уравнений и критерия интерпретируемости факторной структуры как функций от независимых переменных матрицы факторной структуры и значений факторов с ограничениями предлагается осуществлять методом штрафных функций [9]. В качестве метода безусловной оптимизации метода штрафных функций были выбраны метод конфигураций (метод Хука-Дживса) и метод случайного поиска (метод Монте-Карло) [10]. В методе Монте-Карло использовался генератор случайных чисел из встроенной библиотеки Math и функции random языка программирования Java.

Метод штрафов

Даны дважды непрерывно дифференцируемые целевая функция $f(x) = f(x_1, \dots, x_n)$ и функции ограничений $g_j(x) = 0$, $j = 1, \dots, m$; $g_j(x) \leq 0$, $j = m+1, \dots, p$, определяющие множество допустимых решений X .

Требуется найти локальный минимум целевой функции на множестве X , т. е. такую точку $x^* \in X$, что

$$f(x^*) = \min_{x \in X} f(x),$$

$$\text{где } X = \left\{ x \left| \begin{array}{l} g_j(x) = 0, j = 1, \dots, m; m < n \\ g_j(x) \leq 0, j = m+1, \dots, p \end{array} \right. \right\}.$$

Идея метода заключается в сведении задачи на условный минимум к решению последовательности задач поиска безусловного минимума вспомогательной функции:

$$F(x, r^k) = f(x) + P(x, r^k) \rightarrow \min_{x \in R^n},$$

где $P(x, r^k)$ – штрафная функция, r^k – параметр штрафа, задаваемый на каждой k -й итерации. Это связано с возможностью применения эффективных и надёжных методов поиска безусловного экстремума.

Штрафные функции конструируются, исходя из условий:

$$P(x, r^k) = \begin{cases} 0, & \text{when – constraints – are – executed} \\ > 0, & \text{at – non – fulfillment – of – restrictions} \end{cases},$$

причём при невыполнении ограничений и $r^k \rightarrow \infty$, $k \rightarrow \infty$ справедливо $P(x, r^k) \rightarrow \infty$. Чем больше r^k , тем больше штраф за невыполнение ограничений. Как правило, для ограничений типа равенств используется квадратичный штраф, а для ограничений типа неравенств – квадрат срезки:

$$P(x, r^k) = \frac{r^k}{2} \left\{ \sum_{j=1}^m [g_j(x)]^2 + \sum_{j=m+1}^p [g_j^+(x)]^2 \right\},$$

где $g_j^+(x)$ – срезка функции:

$$g_j^+(x) = \max \{0, g_j(x)\} = \begin{cases} g_j(x), & g_j(x) > 0, \\ 0, & g_j(x) \leq 0. \end{cases}$$

Начальная точка поиска задаётся обычно вне множества допустимых решений X . На каждой k -й итерации ищется точка $x^*(r^k)$ минимума вспомогательной функции $F(x, r^k)$ при заданном параметре r^k с помощью одного из методов безусловной минимизации. Полученная точка $x^*(r^k)$ используется в качестве начальной на следующей итерации, выполняемой при возрастающем значении параметра штрафа. При неограниченном возрастании r^k последовательность точек $x^*(r^k)$ стремится к точке условного минимума x^* .

Шаг 1. Задать начальную точку x^0 ; начальное значение параметра штрафа $r^0 > 0$; число $C > 1$ для увеличения параметра; малое число $\varepsilon > 0$ для остановки алгоритма. Положить $k = 0$.

Шаг 2. Составить вспомогательную функцию

$$F(x, r^k) = f(x) + \frac{r^k}{2} \left\{ \sum_{j=1}^m [g_j(x)]^2 + \sum_{j=m+1}^p [g_j^+(x)]^2 \right\}.$$

Шаг 3. Найти точку $x^*(r^k)$ безусловного минимума функции $F(x, r^k)$ по x с помощью какого-либо метода (нулевого, первого или второго порядка):

$$F(x^*(r^k), r^k) = \min_{x \in R^n} F(x, r^k).$$

При этом задать все требуемые выбранным методом параметры. В качестве начальной точки взять x^k . Вычислить $P(x^*(r^k), r^k)$.

Шаг 4. Проверить условие окончания:

а) если $P(x^*(r^k), r^k) \leq \varepsilon$, процесс поиска закончить:

$$x^* = x^*(r^k), f(x^*) = f(x^*(r^k));$$

б) если $P(x^*(r^k), r^k) > \varepsilon$, положить: $r^{k+1} = Cr^k$, $x^{k+1} = x^*(r^k)$, $k = k + 1$ и перейти к шагу 2.

Метод конфигураций

Требуется найти безусловный минимум функции $f(x)$ многих переменных, т. е. найти такую точку $x^* \in R^n$, что $f(x^*) = \min_{x \in R^n} f(x)$.

Шаг 1. Задать начальную точку x^0 , число $\varepsilon > 0$ для остановки алгоритма, начальные величины шагов по координатным направлениям $\Delta_1, \dots, \Delta_n \geq \varepsilon$, ускоряющий множитель $\lambda > 0$, коэффициент уменьшения шага $\alpha > 1$. Положить $y^1 = x^0$, $i = 1, k = 0$.

$$d_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, d_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, d_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Шаг 2. Осуществить исследующий поиск по выбранному координатному направлению:

а) если $f(y^i + \Delta_i d_i) < f(y^i)$, т. е. $f(y_1^i, \dots, y_i^i + \Delta_i, \dots, y_n^i) < f(y_1^i, \dots, y_i^i, \dots, y_n^i)$, шаг считается удачным. В этом случае следует положить $y^{i+1} = y^i + \Delta_i d_i$ и перейти к шагу 3;

б) если в п. «а» шаг неудачен, то делается шаг в противоположном направлении. Если $f(y^i - \Delta_i d_i) < f(y^i)$, т. е. $f(y_1^i, \dots, y_i^i - \Delta_i, \dots, y_n^i) < f(y_1^i, \dots, y_i^i, \dots, y_n^i)$, шаг считается удачным. В этом случае следует положить $y^{i+1} = y^i - \Delta_i d_i$ и перейти к шагу 3;

в) если в пп. «а» и «б» шаги неудачны, положить $y^{i+1} = y^i$.

Шаг 3. Проверить условия:

а) если $i < n$, то положить $i = i + 1$ и перейти к шагу 2 (продолжить исследующий поиск по оставшимся направлениям);

б) если $i = n$, проверить успешность исследующего поиска:

– если $f(y^{n+1}) < f(x^k)$, перейти к шагу 4;

– если $f(y^{n+1}) \geq f(x^k)$, перейти к шагу 5.

Шаг 4. Провести поиск по образцу. Положить $x^{k+1} = y^{n+1}$,

$y^1 = x^{k+1} + \lambda(x^{k+1} - x^k)$, $i = 1$, $k = k + 1$ и перейти к шагу 2.

Шаг 5. Проверить условие окончания:

- а) если все $\Delta_i \leq \varepsilon$, то поиск закончить: $x^* \cong x^k$;
 б) для тех i , для которых $\Delta_i > \varepsilon$, уменьшить величину шага: $\Delta_i = \frac{\Delta_i}{\alpha}$.
 Положить $y^1 = x^k$, $x^{k+1} = x^k$, $k = k + 1$, $i = 1$ и перейти к шагу 2.

3. Численный эксперимент

Численный эксперимент данного научного исследования заключался в тестировании метода оценки параметров структурных уравнений в рамках модели квадратичного факторного анализа на базе 15 биофизических показателей, измеренных у 131-го пациента с артериальной гипертензией начальной стадии:

1. *Вес.*
2. *Индекс массы тела (ИМТ).*
3. *Частота дыхания (ЧД).*
4. *Сегментоядерные нейтрофилы (С).*
5. *Лимфоциты (Л).*
6. *Конечно-систолический размер левого желудочка (КСР).*
7. *Конечно-систолический объём левого желудочка (КСО).*
8. *Конечно-диастолический размер левого желудочка (КДР).*
9. *Конечно-диастолический объём левого желудочка (КДО).*
10. *Ударный объём (УО).*
11. *Минутный объём сердца (МОС).*
12. *Общее периферическое сосудистое сопротивление (ОПСС).*
13. *Индекс Хильдебрандта (ИХ).*
14. *Фракция выброса левого желудочка (ФВ).*
15. *Фракция укорочения левого желудочка (ФУ).*

Предварительно данные были стандартизированы до нулевого среднего и единичной дисперсии. Факторное решение с учётом критерия интерпретируемости представлено в таблице 1. Данная факторная структура подтверждается предыдущими работами [11, 12].

По-прежнему можно говорить о наличии пяти факторов.

1. Гемодинамический фактор, учитывающий структурно-геометрическую модуляцию сердца (*УО, МОС, ОПСС, КСР, КСО, КДР, КДО*).
2. Фактор нагнетательной функции левого желудочка (*ФВ, ФУ*).
3. Фактор, отвечающий за соответствие массы и роста (*вес, ИМТ*).
4. Фактор, характеризующий уровень сложенности работы сердца и лёгких (*ЧД, ИХ*).
5. Иммунологический фактор стресса (*С, Л*).

Наличие дополнительных квадратичных слагаемых в факторных уравнениях привело к улучшению критерия K приблизительно в два раза ($K_{лин}=220$; $K_{квадр}=120$).

Наиболее значимые нагрузки в квадратичной части факторной модели соответствуют следующим парам факторов для исходных переменных.

Знак перед парой факторов соответствует знаку соответствующего значимого коэффициента b_{kij} . Чтобы добиться синергетического эффекта для нор-

Таблица 1. Оценка параметров факторной структуры с учётом критерия интерпретируемости

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>
Вес	-0.2791	0.0984	0.2322	-0.1684	-0.632
ИМТ	-0.2156	0.239	0.1162	-0.1144	-0.6023
ЧД	0.093	0.5298	0.2218	-0.6016	0.1922
С	-0.0367	-0.0637	-0.7378	-0.115	0.0297
Л	0.0497	0.0099	0.7206	0.0596	-0.1085
КСР	-0.6453	0.1778	0.3383	-0.0193	0.2164
КСО	-0.5746	0.093	0.3276	0.0441	0.258
КДР	-0.5698	-0.1579	0.3042	0.4226	0.2668
КДО	-0.5943	-0.1658	0.3033	0.373	0.2701
УО	-0.4668	-0.3725	0.2621	0.5907	0.2401
МОС	-0.5413	-0.3485	0.1823	0.5485	0.2701
ОПСС	0.4748	0.153	-0.2119	-0.3731	-0.2384
ИХ	-0.368	-0.4015	-0.2601	0.573	-0.2227
ФВ	0.3807	-0.6323	-0.2669	0.6098	-0.045
ФУ	0.2959	-0.5831	-0.2515	0.583	0.0968

Таблица 2. Оценка параметров факторной структуры с учётом критерия интерпретируемости

	<i>Пары факторов</i>
Вес	$p_3p_3, -p_2p_3$
ИМТ	$-p_2p_3$
ЧД	
С	$-p_5p_5$
Л	$-p_4p_4, p_5p_5$
КСР	p_2p_4, p_5p_5
КСО	$-p_2p_4, -p_4p_5$
КДР	
КДО	
УО	
МОС	
ОПСС	p_1p_4
ИХ	$-p_2p_2$
ФВ	p_1p_3, p_4p_5
ФУ	

мализации исходных показателей, необходимо одновременно воздействовать на представленные пары факторов. С точки зрения медикаментозной терапии это означает: одновременно пить несколько разных таблеток для нормализации соответствующих факторов, обуславливающих заболевание.

4. Заключение

Предложен метод оценки нелинейных структурных уравнений на базе метода штрафных функций и методов безусловной оптимизации. Все алгоритмы реализованы в виде отдельной программы с графическим интерфейсом для пользователя. Для 15 биофизических показателей артериальной гипертензии начальной стадии и модели квадратичного факторного анализа как частного случая структурного моделирования была получена факторная структура. Полученная факторная структура подтверждается результатами ранних работ. Наличие квадратичных слагаемых в факторных уравнениях привело к улучшению критерия невязок приблизительно в два раза.

ЛИТЕРАТУРА

1. Jereskog K.G., Serbom D. Advances in factor analysis and structural equation models / Edited by Jay Magidson. Cambridge, Mass. : Abt Books, 1979. 242 p.
2. Blau P.M., Duncan O.D., Tyree A. The American Occupational Structure. New York : Wiley and Sons, 1967. 534 p.
3. Blalock H. Theory construction. Englewood Cliffs, New Jersey : Prentice-Hall, 1968. 180 p.
4. Jereskog K.G., Goldberger A.S. Estimation of a model with multiple indicators and multiple causes of a single latent variable // Journal of the American Statistical Association. 1975. V. 70(351). P. 631–639.
5. Bentler P.M. Multivariate analysis with latent variables: Causal modeling // Annual review of psychology. 1980. V. 31(1). P. 419–456.
6. Иберла К. Факторный анализ / Пер. с нем. В.М. Ивановой. М. : Статистика, 1980. 398 с.
7. Харман Г. Современный факторный анализ / Пер. с англ. В.Я. Лумельского. М. : Статистика, 1972. 489 с.
8. Шовин В.А., Гольяпин В.В. Методы вращения факторных структур // Математические структуры и моделирование. 2015. № 2. С. 75–84.
9. Банди Б. Методы оптимизации. Вводный курс. М. : Радио и связь, 1988. 128 с.
10. Кокуев А.Г. Оптимальное управление. Поиск экстремумов многомерных функций. Астрахань : АГТУ, 2011. 34 с.
11. Гольяпин В.В., Шовин В.А. Косоугольная факторная модель артериальной гипертензии первой стадии // Вестник Омского университета. 2010. № 4. С. 120–128.
12. Шовин В.А. Конфирматорная факторная модель артериальной гипертензии // Компьютерные исследования и моделирование. 2012. Т. 4, № 4. С. 885–894.

NONLINEAR STRUCTURAL EQUATIONS AND QUADRATIC FACTOR ANALYSIS

V.A. Shovin

Scientist Researcher, e-mail: v.shovin@mail.ru

Sobolev Institute of Mathematics of Siberian Branch of the Russian Academy of Sciences
(Omsk Branch), Omsk, Russia

Abstract. The role of the article is to develop a new algorithm (method) of factor analysis, which includes an advanced model of the relationship of factors and initial variables, when the model takes into account not only linear terms, but also quadratic ones, i.e. decomposition to the second power. Such a model belongs to the class of nonlinear structural models. The calculation algorithm makes it possible to build a refinement model, to investigate the influence of the nonlinear component in the factor influence. Also, an appropriate program with a graphical user interface was developed that can carry out a quadratic factor analysis of new data on the loading of new data. This article is both applied and fundamental. The program and the corresponding computational algorithm were tested on real data, which allowed us to compare the previous calculations for these data for linear factor analysis with the quadratic one proposed in the paper. Estimates of the parameters of nonlinear structural equations was implemented by numerical methods for nonlinear optimization with conditions. As a criterion for the optimal parameters of structural equation was selected to minimize the residual of structural equation model. The results of the testing of this method for the quadratic factor analysis model of hypertension was in accordance with other methods of factor modeling. This method of estimates of the parameters of structural equations was implemented as software.

Keywords: hypertension, factor analysis, structural equations.

Дата поступления в редакцию: 20.02.2018

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТА ПО ОБНАРУЖЕНИЮ БЕСПРОВОДНОЙ ТОЧКИ ДОСТУПА МОДИФИЦИРОВАННЫМ МЕТОДОМ ТРИЛАТЕРАЦИИ

А.А. Кондюрина

магистрант, e-mail: a.kondyurina@gmail.com

Д.Н. Лавров

к.т.н., доцент, e-mail: lavrov@omsu.ru

Омский государственный университет им. Ф.М. Достоевского, Омск, Россия

Аннотация. В работе представлены результаты проведения эксперимента по обнаружению беспроводной точки доступа в открытом пространстве на основе ранее построенной модели.

Ключевые слова: позиционирование, обнаружение беспроводной точки доступа, трилатерация, тест Колмогорова-Смирнова.

В статье [1] была предложена математическая модель получения оценки координат по измерениям мощности излучения точки беспроводного доступа на основе трилатерации. С помощью компьютерного эксперимента [2] на основе этой модели была найдена положительная корреляция точности и количества наблюдателей в системе.

В рамках эксперимента принимается ряд допущений.

1. Несанкционированно установленная точка доступа (с поддержкой стандартов 802.11) стационарна, не меняет своего положения в пространстве.
2. Не учитываются препятствия для распространения сигнала (перегородки, рельеф местности и т. д.).
3. Наблюдатель перемещается по заранее известной траектории.
4. Не учитывается характер сигналов; принимается, что излучатель — это изотропная антенна.

Несмотря на то, что примерная траектория перемещения устройства наблюдения известна, необходимо её уточнять, чтобы обеспечивать точность определения координат точки измерения мощности сигнала. Для этого предлагается использовать фильтр Калмана с последующим сглаживанием на закреплённом интервале для ускорений, получаемых с акселерометра. Эта система коррекции перемещения описана в работе [3].

На предварительном этапе необходимо определить реальные шумы, воздействующие на систему, в данном случае это помехи, воздействующие на акселерометр мобильного устройства.

Был проведён эксперимент по определению шумов акселерометра на отдельно взятом мобильном устройстве в условиях покоя (ускорения по осям x , y , z стремятся к 0, вектор гравитации не учитывается). Для полученных

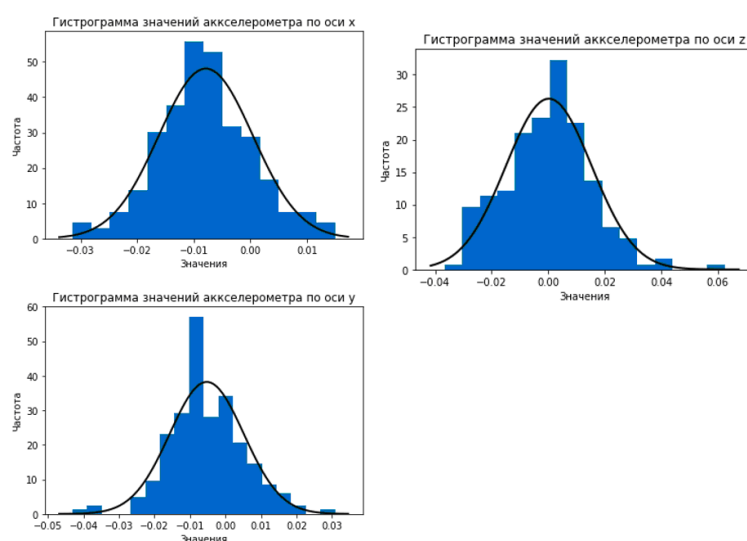


Рис. 1. Гистограммы распределения ошибок показаний ускорения акселерометра мобильного устройства по осям

значений шумов были построены гистограммы по осям и применён критерий Колмогорова–Смирнова для проверки, что шумы системы подчиняются нормальному распределению.

Построенные гистограммы лишь условно показывают приближения распределений величин к нормальному (рис. 1).

В качестве нулевой гипотезы H_0 для теста Колмогорова–Смирнова было выбрано утверждение о том, что распределение ошибок измерения показаний акселерометра подчиняется закону Гаусса.

В результате на выборке были получены значения KS статистики, равной 0.055190481309094375, и достигаемого уровня значимости или пи-величины (p -value) — 0.92089092791638738. Если KS статистика мала или p -value велико, то мы не можем однозначно опровергнуть гипотезу H_0 .

Кроме того, при выбранном уровне значимости $\alpha = 0.05$ полученная пи-величина больше α , что позволяет сделать заключение о применимости предложенной модели и перейти к дальнейшей программной реализации.

Эксперимент по обнаружению точки доступа проводился на открытой территории с примерной длиной каждой из сторон 50 метров, стены и иные препятствия для распространения сигнала отсутствовали.

Устройство наблюдения: смартфон LG Nexus 5 (модель D821). Датчик движения: InvenSense MPU-6515. Датчик беспроводной связи: Broadcom BCM4339.

Искомое беспроводное устройство: смартфон Apple iPhone 7 (модель A1778) с включённым режимом модема (hotspot). Мощность сигнала источника, измеренная на расстоянии $d_0 = 1$ м от устройства (P0): -34 дБм.

Человек с мобильным устройством в руках, выполняющий роль наблюдателя, двигался по траектории сначала вдоль оси OX (от координаты (0, 0) до (50, 0)), затем вдоль оси OY со смещением по OX (от координаты (50, 0) до (50,

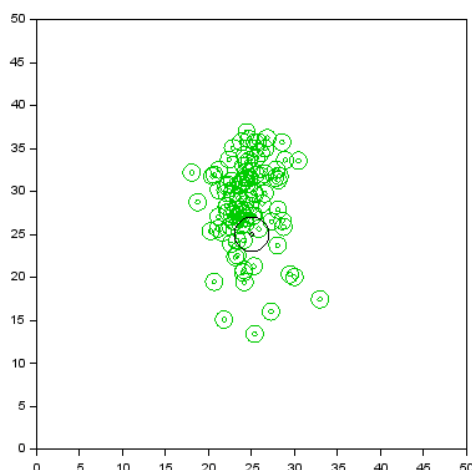


Рис. 2. Оценки местоположения точки доступа по вычисленным координатам наблюдателя и соотнесенным значениям мощности (искомая точка доступа — окружность в центре (25, 25), окружности меньших диаметров — оценки координат)

50)). Точка доступа располагалась в центре (25, 25).

Были получены следующие результаты (рис. 2). На примере для центрального расположения источника сигнала заметно, что оценки координат достаточно сильно разнесло.

Так как вычисленные оценки координат перемещения наблюдателя не показало серьёзных отклонений, возникло предположение, что причина подобного разброса заключается в аппаратной реализации модуля сканирования сетей.

Из-за специфики доступа к сканированию сетей [4] на мобильном устройстве нет возможности получать актуальное значение мощности конкретной точки через очень небольшие интервалы времени.

Также было замечено, что на равноудалённых точках уровень сигнала может различаться на 5–10 дБм, такое поведение приемлемо для стандартного сканирования сетей с целью определения грубой оценки приближённости антенны к наблюдателю. Однако в рамках решаемой задачи такие перепады сильно ухудшают результат.

ЛИТЕРАТУРА

1. Вишнякова О.А., Лавров Д.Н., Лаврова С.Ю. Математическая модель обнаружения точки беспроводного доступа по измерениям мощности излучения разнесёнными наблюдателями // Математические структуры и моделирование. 2013. № 2(28). С. 49–59.
2. Лавров Д.Н., Вишнякова О.А., Дудяк Е.И., Лаврова С.Ю. Компьютерное моделирование оценивания координат точки беспроводного доступа по измерениям мощности принимаемых сигналов // Математические структуры и моделирование. 2014. № 2(30). С. 62–76.

3. Лавров Д.Н., Кондюрина, А.А. Определение траектории движения приёмника системы обнаружения несанкционированно установленной точки беспроводного доступа // Математические структуры и моделирование. 2017. № 1(41). С. 75–87.
4. Android XRef Nougat 7.0.0_r1. Cross Reference: WifiStateMachine.java URL: http://androidxref.com/7.0.0_r1/xref/frameworks/opt/net/wifi/service/java/com/android/server/wifi/WifiStateMachine.java#1595 (дата обращения: 13.05.2018).

RESULTS OF THE WIRELESS ACCESS POINT DETECTION EXPERIMENT BY MODIFIED TRILATERATION METHOD

A.A. Kondyurina

Student, e-mail: a.kondyurina@gmail.com

D.N. Lavrov

Ph.D.(Eng.), Associate Professor, e-mail: lavrov@omsu.ru

Dostoevsky Omsk State University, Omsk, Russia

Abstract. The research paper presents the results of an experiment to detect wireless access point in the open space based on the early suggested model.

Keywords: positioning, wireless access point detection, trilateration, Kolmogorov-Smirnov test.

Дата поступления в редакцию: 21.05.2018

WHY UNDER STRESS POSITIVE REINFORCEMENT IS MORE EFFECTIVE? WHY OPTIMISTS STUDY BETTER? WHY PEOPLE BECOME RESTLESS? SIMPLE UTILITY-BASED EXPLANATIONS

Francisco Zapata

Ph.D. (Phys.-Math.), Instructor, e-mail: fazg74@gmail.com

Olga Kosheleva

Ph.D. (Phys.-Math.), Associate Professor, e-mail: olgak@utep.edu

Vladik Kreinovich

Ph.D. (Phys.-Math.), Professor, e-mail: vladik@utep.edu

University of Texas at El Paso, El Paso, Texas 79968, USA

Abstract. In this paper, we use the utility-based approach to decision making to provide simple answers to the following three questions: Why under stress positive reinforcement is more effective? Why optimists study better? Why people become restless?

Keywords: utility theory, positive vs. negative reinforcement, optimists vs. pessimists.

1. Why Under Stress Positive Reinforcement Is More Effective?

Phenomenon. To encourage a person to do something, we can use both positive reinforcement — when we reward a person for doing this, and negative reinforcement — when we penalize a person for not doing the task. Both approaches have their strengths and limitations.

It has been observed that in stress situations, when a person's mood is negative, the relative strength of positive reinforcement increases; see, e.g., [11, 21]. Why?

Let us formulate this problem in precise terms. In traditional decision theory, human preferences are described by *utilities*; see, e.g., [2, 3, 14, 17, 19]. A utility is defined as follows: we select a very good situation A_1 and a very bad situation A_0 and then compare each situation A with the lottery $L(p)$ in which:

- we get A_1 with probability p and
- we get A_0 with the remaining probability $1 - p$.

For small p , $L(p)$ is close to A_0 and is, thus, much worse than A : $L(p) < A$. For p close to 1, $L(p)$ is close to A_1 and is, thus, much better than A : $A < L(p)$

There is therefore a threshold value p_0 such that:

- for $p > p_0$, we have $A < L(p)$, while
- for $p < p_0$, we have $L(p) < A$.

This threshold value — for which A is (in this sense) equivalent to $A(p_0)$ — is called the *utility* $u(a)$ of the alternative A .

If $p < p'$, then, of course, $L(p')$ is better than $L(p)$. Thus, among several alternatives, we should select a one for which the utility $u(a)$ is the largest.

It is known that the utility of monetary rewards or losses is approximately proportional to the square root of the amount m of money:

- $u(m) = a_+ \cdot \sqrt{m}$ for $m \geq 0$ and
- $u(m) = -a_- \cdot \sqrt{|m|}$ for $m < 0$,

for some values a_+ and a_- ; see, e.g., [6, 12, 13].

We can measure the relative strength of positive and negative reinforcement by comparing the changes in utility if we add or subtract a certain amount of money m .

If we start with a neutral situation, in which we have no money, then the original utility value is 0. Then, after adding the amount m we get the utility $a_+ \cdot \sqrt{m}$, while after subtracting amount m , we lose the utility amount $a_- \cdot \sqrt{m}$. In this case, the ration of positive-to-negative reinforcement effects is

$$\frac{a_+ \cdot \sqrt{m}}{a_- \cdot \sqrt{m}} = \frac{a_+}{a_-}. \quad (1)$$

What if we start with a stressful situation, in which the initial amount of money is small but negative: $-m_0 < 0$? In this case, the initial value of the utility is $a_- \cdot \sqrt{m_0}$. After adding m , we get $m - m_0$, with the utility $a_+ \cdot \sqrt{m - m_0}$. Thus, the utility gain is $a_+ \cdot \sqrt{m - m_0} + a_- \cdot \sqrt{m_0}$.

If we subtract the money amount m , then we end up with the negative amount $-(m + m_0)$, whose utility is $-a_- \cdot \sqrt{m + m_0}$. Thus, the loss of utility is the difference $a_- \cdot \sqrt{m + m_0} - a_- \cdot \sqrt{m_0}$.

Thus, the ratio describing the relative strength of possible reinforcement takes the form

$$\frac{a_+ \cdot \sqrt{m - m_0} + a_- \cdot \sqrt{m_0}}{a_- \cdot \sqrt{m + m_0} - a_- \cdot \sqrt{m_0}}. \quad (2)$$

Our explanation. We will show that the ratio (2) is larger than the ratio (1). This explains the empirical fact that under stress, positive reinforcement is more efficient.

Indeed, for small m_0 , by taking the first two terms of the corresponding Taylor series, we get

$$\sqrt{m - m_0} = \sqrt{m} - \frac{1}{2\sqrt{m}} \cdot m_0 + o(m_0).$$

For small m_0 , we have $\sqrt{m_0} \gg m_0$, thus in the first approximation, we can ignore the terms proportional to m_0 and only consider terms proportional to $\sqrt{m_0}$. So, the numerator of the ratio (2) takes the form

$$a_+ \cdot \sqrt{m - m_0} + a_- \cdot \sqrt{m_0} \approx a_+ \cdot \sqrt{m} + a_- \cdot \sqrt{m_0}.$$

Similarly, we have

$$\sqrt{m + m_0} = \sqrt{m} + \frac{1}{2\sqrt{m}} \cdot m_0 + o(m_0)$$

and thus, in the first approximation, the denominator of the formula (2) takes the form

$$a_- \cdot \sqrt{m + m_0} - a_- \cdot \sqrt{m_0} \approx a_- \cdot \sqrt{m} - a_- \cdot \sqrt{m_0}.$$

Thus, in the first approximation, the ratio (2) has the form

$$\frac{a_+ \cdot \sqrt{m} + a_- \cdot \sqrt{m_0}}{a_- \cdot \sqrt{m} - a_- \cdot \sqrt{m_0}}. \quad (3)$$

We can see that, in comparison to the ratio (1), we increased the numerator and decreased the denominator — as a result, the ratio increases.

This is exactly what we wanted to explain.

Auxiliary analysis: beyond explanation. A natural question is: what if instead of considering stress, we consider euphoria, i.e., we consider situations in which we have a positive initial amount of money m_0 . How will this affect the relative strength of positive and negative reinforcements?

In this case, we start with the utility $a_+ \cdot \sqrt{m_0}$. When we add the amount m , we get the utility $a_+ \cdot \sqrt{m + m_0}$, so the increase in utility is equal to

$$a_+ \cdot \sqrt{m + m_0} - a_+ \cdot \sqrt{m_0}.$$

Vice versa, if we take away the amount m , we get the new utility $-a_- \cdot \sqrt{m - m_0}$, so the loss in utility is equal to

$$a_- \cdot \sqrt{m - m_0} + a_+ \cdot \sqrt{m_0}.$$

In this situation, the ratio describing the relative strength of positive and negative reinforcements takes the form

$$\frac{a_+ \cdot \sqrt{m + m_0} - a_+ \cdot \sqrt{m_0}}{a_- \cdot \sqrt{m - m_0} + a_+ \cdot \sqrt{m_0}}. \quad (4)$$

Similarly to the stress case, in the first approximation, the numerator is approximately equal to $a_+ \cdot \sqrt{m} - a_+ \cdot \sqrt{m_0}$, while the denominator is approximately equal to $a_- \cdot \sqrt{m} + a_+ \cdot \sqrt{m_0}$. Thus, in the first approximation, the ratio (4) takes the form

$$\frac{a_+ \cdot \sqrt{m} - a_+ \cdot \sqrt{m_0}}{a_- \cdot \sqrt{m} + a_+ \cdot \sqrt{m_0}}.$$

We can see that, in comparison to the ratio (1), we decreased the numerator and increased the denominator — as a result, the ratio decreases.

Thus, we conclude that in happy situations, negative reinforcements are more efficient than the positive ones (but do not tell that to your bosses :-).

2. Why Optimists Study Better?

Empirical fact. It is a known fact that optimists study better; see, e.g., [22] and references therein.

Let us describe this situation in precise terms. What does optimism mean in precise terms?

According to the traditional decision theory, if for each possible alternative a , we know the probabilities $p_i(a)$ and utilities $u_i(a)$ of different outcomes i , then a rational person should select an alternative a for which the value $u(a) \stackrel{\text{def}}{=} \sum_i p_i(a) \cdot u_i(a)$ is the largest [2, 3, 14, 17, 19]. In such situations, there is only one rational choice, there is no possibility to show optimism or pessimism.

In practice, however, we rarely know the exact probability and the exact utility of different outcomes. Usually, we only know the bounds $\underline{p}_i(a)$, $\bar{p}_i(a)$, $\underline{u}_i(a)$, and $\bar{u}_i(a)$ on possible values of $p_i(a)$ and $u_i(a)$: $\underline{p}_i(a) \leq p_i(a) \leq \bar{p}_i(a)$ and $\underline{u}_i(a) \leq u_i(a) \leq \bar{u}_i(a)$. For different values of $p_i(a)$ and $u_i(a)$ from the corresponding intervals, we get different values of the overall utility $u(a)$. Thus, instead of a single value $u(a)$, we have an *interval* $[\underline{u}(a), \bar{u}(a)]$ of possible values. How should we make decisions if for each alternative a , we know such an interval $[\underline{u}(a), \bar{u}(a)]$?

Reasonable requirements on rationality of a decision maker lead to the following solution (first proposed by the future Nobel Prize winner Leo Hurwicz): we should select a number $\alpha \in [0, 1]$ and select an alternative for which the combination $\alpha \cdot \bar{u}(a) + (1 - \alpha) \cdot \underline{u}(a)$ is the largest possible; see, e.g., [5, 10, 14].

When $\alpha = 1$, this means that when making a decision, we only take into account the most favorable situation, when the utility $u(a)$ attains its largest possible value $\bar{u}(a)$. This is clearly the case of extreme optimism.

When $\alpha = 0$, this means that when making a decision, we only take into account the least favorable situation, when the utility $u(a)$ attains its smallest possible value $\underline{u}(a)$. This is clearly the case of extreme pessimism.

Values α intermediate between 0 and 1 describe realistic decision makers. The larger α , the more the decision maker takes into account the most optimistic scenario and the less he/she takes into account the most pessimistic scenarios. Thus, the value α can serve as a quantitative measure of the decision maker's optimism: the larger α , the more optimistic the decision maker.

This explains why optimistic study better. In education, we invest some efforts now and get rewards in the future. Time for studying is taken from time of having fun: we have less chances to go to a movie, to watch TV, etc. So, in comparison with not studying, this part of the learning process brings less positive utility.

We do study, because we know that there will be a future reward: better knowledge, better job, etc. So, when deciding how much time we dedicate to studying (or whether to study at all), we take into account both the utility decrease now and the potential utility increase in the future.

The decrease now is clear, we thus know the value $u_d < 0$. About the future rewards, we are not 100% certain: now there is a demand for your major, who

knows what will happen four years from now, when we graduate with a degree? Thus, for future rewards, instead of the exact value u_r , we only know the interval of possible values $u_r \in [\underline{u}_r, \bar{u}_r]$. The overall utility therefore takes all possible values from $\underline{u} = u_d + \underline{u}_r$ to $\bar{u} = u_d + \bar{u}_r$. A person with an optimism value α selects to study if the Hurwicz combination $\alpha \cdot \bar{u} + (1 - \alpha) \cdot \underline{u}$ is larger than the value 0 corresponding to not studying.

Here, as one can easily check, the Hurwicz combination is equal to

$$u_d + \alpha \cdot \bar{u}_r + (1 - \alpha) \cdot \underline{u}_r = u_d + \underline{u}_r + \alpha \cdot (\bar{u}_r - \underline{u}_r).$$

This value increases with α . If this value was larger than 0 for some α , it will be still larger than 0 for $\alpha' > \alpha$ — and for $\alpha' > \alpha$, in some situations when the Hurwicz value was negative, it may become positive.

Thus, the larger the level α of a person's optimism, the more there are situations in which this person will start studying. This explains why optimists are better students.

3. Why People Become Restless?

Phenomenon. When a person's salary is increased, this person becomes happy. If a few years pass and the salary remains the same, then, while objectively, the person has the same good life as before, he or she becomes restless, unhappy. Why?

The situation is the same as in the past years, so why is not level of happiness the same?

Towards an explanation. It is known that our utility depends not only on what we have now, it also depends on what we expect in the future: otherwise, we would act without thinking of possible consequences. The expected future values of utility come with some discounting, usually, the exponential discounting, when — just like when you invest money in a bank — the utility T moments in the future gets multiply by β^T for some $\beta < 1$; see, e.g., [1–5, 8–11]. (For the bank, β is 1 minus interest; e.g., if the interest rate is 3%, $\beta = 0.97$.) As a result, if o_t is the utility caused by the current situation at moment t , the actual utility u_t at moment t is equal to

$$u_t = o_t + \alpha \cdot o_{t+1} + \alpha^2 \cdot o_{t+2} + \dots$$

We do not know the future values, we get them by extrapolation, based on the previous several values $o_t, o_{t-1}, \dots$

The simplest possible extrapolation is linear extrapolation which is based on the last two values o_t and o_{t-1} . Here, $o_{t+j} = o_t + j \cdot (o_t - o_{t-1})$. In the year t in which a salary got increased, the difference is positive, so $o_{t+1} > o_t$, $o_{t+2} > o_t$, etc., hence

$$u_t = o_t + \alpha \cdot o_{t+1} + \alpha^2 \cdot o_{t+2} + \dots > o_t + \alpha \cdot o_t + \alpha^2 \cdot o_t + \dots = o_t \cdot (1 + \alpha + \alpha^2 + \dots).$$

A few years later, when $o_t = o_{t-1}$, all extrapolated values are the same: $o_{t+1} = o_{t+2} = \dots = o_t$, thus

$$u_t = o_t + \alpha \cdot o_t + \alpha^2 \cdot o_t + \dots = o_t \cdot (1 + \alpha + \alpha^2 + \dots).$$

We see that the utility in the first year is indeed larger than the utility a few years after — this is exactly what we observe.

Acknowledgments

This work was supported in part by the National Science Foundation grant HRD-1242122 (Cyber-ShARE Center of Excellence).

REFERENCES

1. Critchfield T.S., Kollins S.H. Temporal discounting: basic research and the analysis of socially important behavior // *Journal of Applied Behavior Analysis*. 2001. V. 34. P. 101–122.
2. Fishburn P.C. *Utility Theory for Decision Making*. John Wiley & Sons Inc., New York, 1969.
3. Fishburn P.C. *Nonlinear Preference and Utility Theory*. Baltimore, Maryland : The John Hopkins Press, 1988.
4. Frederick S., Loewenstein G., O'Donoghue T. Time discounting: a critical review // *Journal of Economic Literature*. 2002. V. 40. P. 351–401.
5. Hurwicz L. *Optimality Criteria for Decision Making Under Ignorance*. Cowles Commission Discussion Paper, Statistics, No. 370, 1951.
6. Kahneman D. *Thinking Fast and Slow*. Farrar, Straus, and Girous, New York, 2011.
7. King G.R., Logue A.W., Gleiser D. Probability and delay in reinforcement: an examination of Mazur's equivalence rule // *Behavioural Processes*. 1992. V. 27. P. 125–138.
8. Kirby K.N. Bidding on the future: evidence against normative discounting of delayed rewards // *Journal of Experimental Psychology (General)*. 1997. V. 126. P. 54–70.
9. Konig C.J., Kleinmann M. Deadline rush: a time management phenomenon and its mathematical description. *The Journal of Psychology*. 2005. V. 139, No. 1. P. 33–45.
10. Kreinovich V. Decision making under interval uncertainty (and beyond) // *Human-Centric Decision-Making Models for Social Sciences* / P. Guo, W. Pedrycz (eds.). Springer Verlag, 2014. P. 163–193.
11. Lighthall N.R., Gorlick M.A., Schoeke A., Frank M.J., Mather M. Stress modulates reinforcement learning in younger and older adults // *Psychology of Aging*. 2013. V. 28, No. 1. P. 35–46.
12. Lorkowski J., Kreinovich V. Granularity helps explain seemingly irrational features of human decision making // *Granular Computing and Decision-Making: Interactive and Iterative Approaches* / W. Pedrycz, S.-M. Chen (eds.). Springer Verlag, Cham, Switzerland, 2015. P. 1–31.
13. Lorkowski J., Kreinovich V. *Bounded Rationality in Decision Making Under Uncertainty: Towards Optimal Granularity*. Springer Verlag, Cham, Switzerland, 2018.
14. Luce R.D., Raiffa R. *Games and Decisions: Introduction and Critical Survey*. Dover, New York, 1989.
15. Mazur J.E. An adjustment procedure for studying delayed reinforcement // *Quantitative Analyses of Behavior*. Vol. 5, *The Effect of Delay and Intervening Events* / M.L. Commons, J.E. Mazur, J.A. Nevin, H. Rachlin (eds.). Erlbaum, Hillsdale, 1987.

16. Mazur J.E. Choice, delay, probability, and conditional reinforcement // *Animal Learning Behavior*. 1997. V. 25. P. 131–147.
17. Nguyen H.T., Kosheleva O., Kreinovich V. Decision making beyond Arrow's 'impossibility theorem', with the analysis of effects of collusion and mutual attraction // *International Journal of Intelligent Systems*. 2009. V. 24, No. 1. P. 27–47.
18. Rachlin H., Raineri A., Cross D. Subjective probability and delay // *Journal of the Experimental Analysis of Behavior*. 1991. V. 55. P. 233–244.
19. Raiffa H. *Decision Analysis*. Addison-Wesley, Reading, Massachusetts, 1970.
20. Zapata F., Kosheleva O., Kreinovich V., Dumrongpookaphan T. Do it today or do it tomorrow: empirical non-exponential discounting explained by symmetry ideas // *Proceedings of the International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making IUKM'2018* / V.-N. Huynh, M. Inuiguchi, D.-H. Tran, T. Denoeux (eds.). Hanoi, Vietnam, March 13–15, 2018.
21. Zimmer K., Tough choices // *The Scientist*. 2018. V. 32, No. 3. P. 20–21.
22. Seligman M. *The Optimistic Child: A Proven Program to Safeguard Children Against Depression and Build Lifelong Resilience*. Martin Books, New York, 2007.

**ПОЧЕМУ В УСЛОВИЯХ СТРЕССА ПОЛОЖИТЕЛЬНОЕ ПОДКРЕПЛЕНИЕ
БОЛЕЕ ЭФФЕКТИВНО? ПОЧЕМУ ОПТИМИСТЫ ЛУЧШЕ УЧАТСЯ?
ПОЧЕМУ ЛЮДИ СТАНОВЯТСЯ БЕСПОКОЙНЫМИ? ПРОСТЫЕ
ПОЯСНЕНИЯ НА ОСНОВЕ ПОЛЕЗНОСТИ**

Ф. Запата

к.ф.-м.н., преподаватель, e-mail: fazg74@gmail.com

О. Кошелева

к.ф.-м.н., доцент, e-mail: olgak@utep.edu

В. Крейнович

к.ф.-м.н., профессор, e-mail: vladik@utep.edu

Техасский университет в Эль Пасо, США

Аннотация. В этой статье мы используем подход к принятию решений, основанный на полезности, чтобы дать простые ответы на следующие три вопроса: почему в условиях стресса положительное подкрепление более эффективно? Почему оптимисты лучше учатся? Почему люди становятся беспокойными?

Ключевые слова: теория полезности, положительное и отрицательное подкрепление, оптимисты и пессимисты.

Дата поступления в редакцию: 05.04.2018

ECONOMICS OF COMMITMENT: WHY GIVING AWAY SOME FREEDOM MAKES SENSE

Vladik Kreinovich

Ph.D. (Phys.-Math.), Professor, e-mail: vladik@utep.edu

Olga Kosheleva

Ph.D. (Phys.-Math.), Associate Professor, e-mail: olgak@utep.edu

Mahdokht Afravi

Doctoral Student, e-mail: mafravi@miners.utep.edu

Genesis Bejarano

Graduate Student, e-mail: gbejarano@miners.utep.edu

Marisol Chacon

Graduate Student, e-mail: mchacon3@miners.utep.edu

University of Texas at El Paso, El Paso, Texas 79968, USA

Abstract. In general, the more freedom we have, the better choices we can make, and thus, the better possible economic outcomes. However, in practice, people often artificially restrict their future options by making a commitment. At first glance, commitments make no economic sense, and so their ubiquity seems puzzling. Our more detailed analysis shows that commitment often makes perfect economic sense: namely, it is related to the way we take future gains and losses into account. With the traditionally assumed exponential discounting, commitment indeed makes no economic sense, but with the practically observed hyperbolic discounting, commitment is indeed often economically beneficial.

Keywords: commitment, economics, discounting, exponential discounting, hyperbolic discounting.

1. Formulation of the Problem

Case study: a simplified description. This study started with a situation which, at first glance, sounds very reasonable. To get to the gist of it and to avoid irrelevant details, we will describe a simplified version of this situation.

There are several Jewish congregations in El Paso. To coordinate their activities, these organizations form the Jewish Federation. To support its activities, each congregation collects money from its congregants. A certain agreed-upon portion of this money is given to the Federation.

The Federation's Board — which consists of representatives of different congregations — decides how to spend this money. Congregations submit proposals, and the Board decides which proposals to grant and to what extent. In making these decisions, the Board tries its best to make sure that the overall amount of grants

given to each congregation is proportional to the amount that the congregation donated to the Federation's budget.

From the economic viewpoint, this is a problem. At first glance, this sounds like a reasonable and fair procedure. However, if one thinks about it from the economic viewpoint, this procedure does not seem to make much sense.

Each congregation pays a certain amount of the money to the Federation, and then gets approximately the same amount of money back. So why not spend it directly? Why do we need this complicated process in the first place? Why not reserve the Federation funding scheme only for emergencies or for unusual situations?

Not only this scheme seems to make no economic sense, it seems to make the situation worse for each congregation: when the original sum of money was the congregation's money, the congregation could spend it on whatever project it could think about (and to which the congregants would agree). However, now that the same amount of money has come through the Federation process, it comes as Federation grants. The congregation is no longer free to spend this money any way it wants: it can only spend it on the project that it originally proposed to the Federation.

So, the congregation got its own money back, but it lost freedom in spending this money: it cannot easily change its mind and spend this money on something else, it has to stick to the original plan.

So why would the congregations go to all this process? Why not keep most of the money to itself and spend it any way the congregation wants?

Commonsense explanation and other cases of commitment. A commonsense explanation is that the Federation scheme brings commitment. If a congregation simply wants to sponsor some future event — e.g., a concert or a conference, it can decide to allocate some funds for this event. But there is always a need of day-by-day funding: the building become shabby and may need maintenance, the air conditioning equipment wears out and does not work as well anymore, there is a desire to decorate the building, etc. When there is a minor project like this worth doing, and there is a pool of money available for some future event — it is tempting to take money from this future-event fund. As a result, often, the future-event funding drastically decreases — and the event never happens.

On the other hand, by signing to the Federation-based scheme, the congregation makes a commitment to spend the allocated money exclusively on the future event — this avoiding the temptation of spending this amount to pay for day-by-day operations.

Such a commitment procedures are ubiquitous. To avoid temptation, people make pledges to stick to their diets, to go to the gym, to save money for retirement — often agreeing on penalties if they violate this commitment. For example, a known advice for a politically involved person to stick to their diet or to their exercise routine is to sign a check to an opposite political organization and give it to a trusted friend (or, better, a lawyer) with an instruction to send it when the check-signer stops dieting or exercising.

Remaining problem. From the commonsense viewpoint, the above explanation may be somewhat convincing, but from the economic viewpoint, it still remains a mystery.

There is a lot of evidence that freedom of choice is good for a person (see, e.g., [6, 7] and references therein), so why would a person artificially restrict his/her freedom of choice?

What we do in this paper. In this paper, we provide an economic explanation for the benefits of commitment.

2. Our Explanation

Main idea: let us take discounting into account. Our explanation is based on the fact that when people make decisions, they take into account not only the amount of money (or whatever good) that they get at the given moment of time, they also take into account the future gains or losses — of course, with some discounting, so that \$1 in the future is valued less than \$1 right now.

Formally, let $m(t)$ denote the amount of money (or other goods) that a person (or an organization) gets at moment t . If the person only took into account the current gains — and ignored possible consequences — then a natural idea would be to select an action for which the current gain $m(t)$ is the largest possible. Let $D(t_0)$ denote the discounting of an event t_0 moments of time in the future, i.e., the amount of money now which is equivalent to \$1 at time t_0 in the future. In these terms, the overall gain from a given action is determined by the formula

$$m = m(t) + D(1) \cdot m(t+1) + D(2) \cdot m(t+2) + \dots, \quad (1)$$

and we select an action for which this value m is the largest possible.

What discounting people use? If we place \$1 in the bank, where it will grow interest at a rate $r\%$, then after t_0 years, this amount will increase to a larger amount q^{t_0} , where $q \stackrel{\text{def}}{=} 1 + \frac{r}{100}$. To get \$1 at moment t_0 , we thus need to invest the amount $\frac{1}{q^{t_0}} = q^{-t_0}$. From this viewpoint, \$1 at moment t_0 is equivalent to q^{-t_0} dollars right now. It therefore seems reasonable to take $D(t_0) = q^{-t_0}$.

This *exponential discounting* was indeed the original idea of decision theorists about how people make decisions. However, since the exponential function decreases to practically 0 fast, it would mean that people, in effect, ignore future consequences when making decisions. Such behavior happens — e.g., when a criminal robs a bank without thinking of future jail time or a young man takes heavy drugs without thinking of long-term damage to his brain — but luckily, such behavior is rare. Empirical data shows that when people make decisions, they use the formula

$$D(t_0) = \frac{1}{1 + k \cdot t_0} \quad (2)$$

known as *hyperbolic discounting*; see, e.g., [1–5, 8–11].

Hyperbolic discounting explains economic efficiency of commitment. Let us show that taking hyperbolic discounting into account explains the economic advantages of commitment.

For simplicity, let us start counting time from the moment when we make the original decision, so that the original decision corresponds to moment $t = 0$. At this moment, we plan an event at moment t_E . We can gauge the positive consequences of this event by estimating the amount m_E that we were willing to pay to have this event right now — if we had the corresponding amount available right now.

Before this event, at some moment t_T , there is a *temptation* (e.g., desire to decorate the building). The negative effect of ignoring this temptation can also be gauged by the corresponding monetary amount; let us denote it by m_T . So, if we decide to keep the event and ignore the temptation, we thus get the equivalent amount

$$m = -D(t_T) \cdot m_T + D(t_E) \cdot m_E. \quad (3)$$

If we do not sponsor the event, then we do not get any gain at moment t_E but we also do not suffer any negative consequences at the moment t_T , so our overall effect is 0.

So, if the amount (3) is greater than 0, we decide to sponsor the event. For hyperbolic discounting (2), the inequality $m > 0$ means that

$$-m_T \cdot \frac{1}{1 + k \cdot t_T} + m_E \cdot \frac{1}{1 + k \cdot t_E} > 0, \quad (4)$$

i.e., equivalently, that

$$m_E \cdot (1 + k \cdot t_T) - m_T \cdot (1 + k \cdot t_E) > 0. \quad (5)$$

Formula (4) implies that $m_E > m_T$: otherwise, even with a larger discounting, the gain m_E would not outweigh the losses m_T .

If at this moment $t = 0$, we make a commitment, we will sponsor the desired event. But suppose that we did not make this commitment. Then, in principle, when the time t_T of the temptation arrives, we can reconsider. At this time, we still prefer to hold the event if

$$-m_T + m_E \cdot \frac{1}{1 + k \cdot (t_E - t_T)} > 0,$$

i.e., if $m_E > m_T \cdot (1 + k \cdot (t_E - t_T))$. If $m_E < m_T \cdot (1 + k \cdot (t_E - t_T))$, we will be tempted to reconsider — and for a large difference $t_E - t_T$, the product $m_T \cdot (1 + k \cdot (t_E - t_T))$ is also large, so this inequality is quite possible. Thus, if we do not make a commitment, it is highly possible that we will reconsider and the desired event will not happen.

Thus, commitment indeed makes perfect economic sense.

Comment. Interestingly, the above argument only holds for the hyperbolic discounting. For exponential discounting, if the initial inequality

$$-q^{-t_T} \cdot m_T + q^{-t_E} \cdot m_E > 0 \quad (6)$$

is satisfied, then at every future moment t , multiplying both sides of inequality (6) by q^t , we get the new inequality

$$-q^{-(t_T-t)} \cdot m_T + q^{-(t_E-t)} \cdot m_E > 0$$

meaning that we should still be committed to holding the event.

Thus, the economics of commitment is based on the fact that our discounting is hyperbolic: for exponential discounting, there is no economic need for commitment.

Acknowledgments

This work was supported in part by the US National Science Foundation grant HRD-1242122.

REFERENCES

1. Critchfield T.S., Kollins S.H. Temporal discounting: basic research and the analysis of socially important behavior // *Journal of Applied Behavior Analysis*. 2001. V. 34. P. 101–122.
2. Frederick S., Loewenstein G., O'Donoghue T. Time discounting: a critical review // *Journal of Economic Literature*. 2002, Vol. 40, pp. 351–401.
3. King G.R., Logue A.W., Gleiser D. Probability and delay in reinforcement: an examination of Mazur's equivalence rule // *Behavioural Processes*. 1992. V. 27. P. 125–138.
4. Kirby K.N. Bidding on the future: evidence against normative discounting of delayed rewards // *Journal of Experimental Psychology (General)*. 1997. V. 126. P. 54–70.
5. König C.J., Kleinmann M. Deadline rush: a time management phenomenon and its mathematical description // *The Journal of Psychology*. 2005. V. 139, No. 1. P. 33–45.
6. Kreinovich V. S. Maslov's iterative method: 15 years later (freedom of choice, neural networks, numerical optimization, uncertainty reasoning, and chemical computing // *Problems of Reducing the Exhaustive Search* / V. Kreinovich and G. Mints (eds.). American Mathematical Society, Providence, RI, 1997. P. 175–189.
7. Kreinovich V., Fuentes O. High-concentration chemical computing techniques for solving hard-to-solve problems, and their relation to numerical optimization, neural computing, reasoning under uncertainty, and freedom of choice // *Molecular and Supramolecular Information Processing: From Molecular Switches to Logical Systems* / E. Katz (ed.). Wiley-VCH, Weinheim, Germany, 2012. P. 209–235.
8. Mazur J.E. An adjustment procedure for studying delayed reinforcement // *Quantitative Analyses of Behavior*. Vol. 5, *The Effect of Delay and Intervening Events* / M.L. Commons, J.E. Mazur, J.A. Nevin, H. Rachlin (eds.). Erlbaum, Hillsdale, 1987.
9. Mazur J.E. Choice, delay, probability, and conditional reinforcement // *Animal Learning Behavior*. 1997. V. 25. P. 131–147.
10. Rachlin H., Raineri A., Cross D. Subjective probability and delay // *Journal of the Experimental Analysis of Behavior*. 1991. V. 55. P. 233–244.
11. Zapata F., Kosheleva O., Kreinovich V., Dumrongpookaphan T. Do it today or do it tomorrow: empirical non-exponential discounting explained by symmetry ideas // *Proceedings of the International Symposium on Integrated Uncertainty in Knowledge*

Modelling and Decision Making IUKM'2018 / V.-N. Huynh, M. Inuiguchi, D.-H. Tran, T. Denoeux (eds.). Hanoi, Vietnam, March 13–15, 2018.

ЭКОНОМИКА ОБЯЗАТЕЛЬСТВ: ПОЧЕМУ ИМЕЕТ СМЫСЛ ЗАРАНЕЕ НЕМНОГО ОГРАНИЧИВАТЬ СВОЮ СВОБОДУ ДЕЙСТВИЙ

В. Крейнович

к.ф.-м.н., профессор, e-mail: vladik@utep.edu

О. Кошелева

к.ф.-м.н., доцент, e-mail: olgak@utep.edu

М. Афрави

докторант, e-mail: mafravi@miners.utep.edu

Дж. Бехарано

аспирант, e-mail: gbejarano@miners.utep.edu

М. Чакон

аспирант, e-mail: mchacon3@miners.utep.edu

Техасский университет в Эль Пасо, США

Аннотация. В целом, чем больше свободы у нас есть, тем лучше выбор, который мы можем сделать, и, следовательно, лучше возможные экономические результаты. Однако, на практике люди часто искусственно ограничивают свои будущие варианты, беря на себя некоторое обязательство. На первый взгляд, обязательства не имеют экономического смысла, и поэтому их повсеместность кажется озадачивающей. Наш более подробный анализ показывает, что обязательство часто имеет прекрасный экономический смысл, а именно: оно связано с тем, как мы учитываем будущие выгоды и потери. При традиционно предполагаемом экспоненциальном дисконтировании обязательство действительно не имеет никакого экономического смысла, но при практически наблюдаемом гиперболическом дисконтировании обязательство действительно часто является экономически выгодным.

Ключевые слова: обязательство, экономика, дисконтирование, экспоненциальная дисконтирование, гиперболический дисконтирование.

Дата поступления в редакцию: 10.04.2018

АЛГОРИТМ ФОРМИРОВАНИЯ ОБУЧАЮЩЕГО МНОЖЕСТВА ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ ДЛЯ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЯ

С.В. Белим

профессор, д.ф.-м.н., профессор кафедры информационной безопасности,
e-mail: sbelim@mail.ru

С.Б. Ларионов

аспирант кафедры информационной безопасности, e-mail: me@stas-larionov.ru

Омский государственный университет им. Ф.М. Достоевского, Омск, Россия

Аннотация. В статье предложен алгоритм формирования обучающего множества для искусственной нейронной сети в задаче сегментации изображений. Отличительной особенностью является использование только одного сегментируемого изображения. Сегментация выполняется с помощью трёх-слойного персептрона. Используется метод выращивания областей. Нейронная сеть используется для принятия решения о принадлежности пикселя к формируемому сегменту. Для формирования обучающего множества используется импульсный шум. Пиксели, изменённые шумом, не принадлежат ни одному сегменту. Проведён компьютерный эксперимент в автоматическом и интерактивном режиме.

Ключевые слова: сегментация изображений, искусственные нейронные сети, импульсный шум.

Введение

Сегментация изображений представляет собой один из вариантов задачи кластеризации множества пикселей по признаку цветовой близости. Следует отметить, что задача сегментации не имеет точного решения и эффективность решения принято оценивать на основе сравнения с результатами, полученными экспертами в ручном режиме. То есть эффективность сегментации не может быть однозначной. Одним из подходов, активно развивающимся в последние годы, является использование искусственных нейронных сетей. Однако данный подход сталкивается с проблемой формирования обучающего множества. Если есть серия аналогичных изображений, например, с медицинской информацией, то проблема решается достаточно просто. Если ставится задача сегментировать одно отдельно взятое изображение, не имеющее аналогов, то традиционные подходы сталкиваются с неразрешимыми трудностями.

Для решения задачи сегментации изображения использовались различные искусственные нейронные сети. Были построены алгоритмы на основе много-

слойного персептрона [1], а также его модификации, использующие перекрёстную энтропию [2], генетические алгоритмы [3,4], выращивание областей [5,6], минимальные отношения различия [7], вейвлет-разложения исходного изображения [8]. Также многослойный персептрон использовался совместно с другими алгоритмами сегментации, например, методом k -средних [9]. Также был разработан ряд алгоритмов на основе самоорганизующихся карт Кохонена [10–12]. Также как и для многослойного персептрона были предложены алгоритмы, совмещающие карты Кохонена с другими алгоритмами сегментации – гибридным генетическим алгоритмом [13], методом k -средних [14], выращиванием областей [15].

Все перечисленные выше алгоритмы носят специализированный характер и ориентированы на изображения определённого типа. Такая специализация необходима для формирования обучающего множества. Нейронные сети, обученные на одном типе изображений, не работают для других изображений.

В данной статье предложен алгоритм формирования обучающего множества в задаче сегментации, ориентированный на одно конкретное изображение и не требующий дополнительного набора схожих изображений.

1. Постановка задачи и метод её решения

Алгоритм сегментации изображения будет основываться на методе выращивания областей, который хорошо зарекомендовал себя при использовании графового представления изображения [16,17]. Выращивание сегмента начинается из одной точки, выбранной случайно на необработанной части изображения. В начальный момент времени всё изображение считается необработанным. Все точки, присоединённые к одному из сегментов, считаются обработанной частью изображения. Выращивание сегмента происходит путём присоединения к нему новых точек. Алгоритм осуществляет обход граничных точек уже сформированной части сегмента и попарно обрабатывает их с соседними точками, не входящими в сегмент. Для каждой такой пары принимается решение о присоединении или неприсоединении к сегменту. Присоединяются к сегменту точки по цветовым характеристикам схожие с уже входящими в него точками.

Для каждого сегмента используется уникальная метка, представленная натуральным числом. Необработанная часть изображения маркируется нулевым значением. Таким образом, на каждом этапе работы алгоритма всем точкам изображения сопоставлены некоторые метки. В начальный момент времени все точки имеют нулевую метку. По окончании работы алгоритма всем точкам сопоставлены ненулевые метки.

Для принятия решения о схожести двух точек был использован трёхслойный персептрон. Точки сравнивались по их цветовым координатам в модели RGB. Так как сравнение происходит только для двух точек, то входной слой персептрона имеет шесть нейронов, соответствующих цветовым координатам точек. Выходной слой содержит два нейрона, соответствующих положительному и отрицательному решению о схожести точек. В рамках компьютерного эксперимента было определено, что эффективность работы алгоритма растёт с

увеличением количества нейронов в скрытом слое вплоть до значения 50, после чего остаётся неизменным. Поэтому в дальнейшем использовался трёхслойный персептрон с пятьюдесятью нейронами в скрытом слое.

Будем рассматривать изображение размером $N \times M$ пикселей, которое в модели RGB может быть представлено тремя матрицами $\|R_{ij}\|$, $\|G_{ij}\|$, $\|B_{ij}\|$ ($i = 1, \dots, N$; $j = 1, \dots, M$), соответствующими разным цветовым компонентам. Для пикселя с координатами (i, j) введём обозначение через v_{ij} . Цветовые составляющие пикселя v_{ij} будем рассматривать как его поля данных:

$$v_{ij}.r = R_{ij}, \quad v_{ij}.g = G_{ij}, \quad v_{ij}.b = B_{ij}.$$

Для матрицы меток введём обозначение c_{ij} ($i = 1, \dots, N$; $j = 1, \dots, M$). Для метки отдельного пикселя также будем использовать представление в виде поля данных: $v_{ij}.c = C_{ij}$.

Рассмотрим алгоритм выращивания сегмента. Будем считать, что к данному моменту выделено k сегментов. Из множества пикселей с нулевой меткой случайным образом выбираем один, который обозначим v_{ij} . Для данного пикселя изменяем метку $v_{ij}.c = k + 1$. Далее запускаем рекурсивный алгоритм.

1. Пусть выращиваемый сегмент содержит g пикселей $\{v_1, \dots, v_r\}$.
2. Перебираем ближайших соседей, имеющих нулевую метку пикселей входящих в сегмент.
3. Формируем входной вектор нейросети из цветовых координат точки сегмента v и её ближайшего соседа v' , не принадлежащего сегменту

$$(v.r, v.g, v.b, v'.r, v'.g, v'.b).$$

4. Вычисляем выходное значение нейросети. При положительном решении изменяем метку пикселя v' : $v'.c = k + 1$. При отрицательном решении переходим к следующему соседнему пикселю, не входящему в сегмент

5. Алгоритм по формированию сегмента останавливается, если нейросеть принимает отрицательное решение для всех ближайших соседей.

Сегментация изображения завершается, если не остаётся ни одного пикселя с нулевой меткой. Алгоритм будет иметь линейную трудоёмкость, так как осуществляется проход по всем пикселям, каждый из которых имеет не более восьми ближайших соседей.

Эффективность алгоритма существенно зависит от обучения нейронной сети, которое определяется обучающей выборкой. Будем формировать обучающее множество для отдельно взятого изображения — искусственно сгенерированный импульсный шум. Используем генератор псевдослучайной последовательности с равномерным распределением. На основе этого генератора сформируем новые цветовые координаты для p % пикселей исходного изображения. Причём координаты изменяемых пикселей будем выбирать произвольно, а новый цвет так, чтобы он значительно отличался от исходного. Будем считать, что вся палитра содержит m цветов. Если в исходном изображении для изменяемого пикселя с координатами (i, j) одна из цветовых координат имела значение $m_{ij} > m/2$, то новый цвет выбираем из интервала $[0, m/2)$. Для случая

$m_{ij} \leq m/2$ новый цвет выбирается из интервала $(m/2, m - 1]$. В результате такого выбора можно считать, что изменённый пиксель не может быть объединён в один сегмент с ближайшими соседями. Изменённые пиксели и играют роль обучающей выборки. Как показал компьютерный эксперимент, необходимо выбирать значение $p \leq 10\%$. Для формирования обучающего множества большего объёма необходимо повторить процедуру формирования импульсного шума на исходном изображении несколько раз. Обучение нейросети выполнялось методом обратного распространения ошибки. Основным недостатком такого подхода к формированию обучающего множества является то, что мы получаем набор только отрицательных решений.

2. Компьютерный эксперимент

В рамках компьютерного эксперимента осуществлялась сегментация цветных изображений с глубиной палитры каждого цвета $m = 256$. Повреждённые пиксели составляли 10 % и формировались 100 раз. Алгоритм тестировался как в автоматическом, так и в интерактивном режиме. В автоматическом режиме определялись все сегменты изображения. В интерактивном режиме осуществлялся поиск одного сегмента, содержащего точку, определяемую пользователем.

Пример работы алгоритма в автоматическом режиме представлен на рисунке 1, на котором приведены контуры сегментов. На рисунке 2 представлен результат работы алгоритма в интерактивном режиме.

Заключение

Не смотря на имеющиеся недостатки в формировании обучающего множества, предложенный алгоритм позволяет выполнять сегментацию заданного изображения с помощью нейронной сети без привлечения других близких по структуре изображений. Компьютерный эксперимент показал, что предложенный алгоритм наиболее эффективен в случае чётких контуров и сегментов большого размера. Следует отметить, что данное свойство присуще всем алгоритмам сегментации. Однако необходимо отметить устойчивость предложенного нами алгоритма сегментации к импульсным шумам, которые легко определяются обученной нейронной сетью как сегменты размером в один пиксель. В дальнейшем данный подход может быть использован с другими более эффективными нейронными сетями.

Также предложенный алгоритм в силу линейной трудоёмкости обладает хорошим быстродействием. Основное время тратится на формирование обучающего множества и обучение искусственной нейронной сети. Процесс сегментации на уже обученной нейросети выполняется значительно быстрее. Общее время обработки изображения размером 256×256 пикселей составляет порядка 40 секунд.

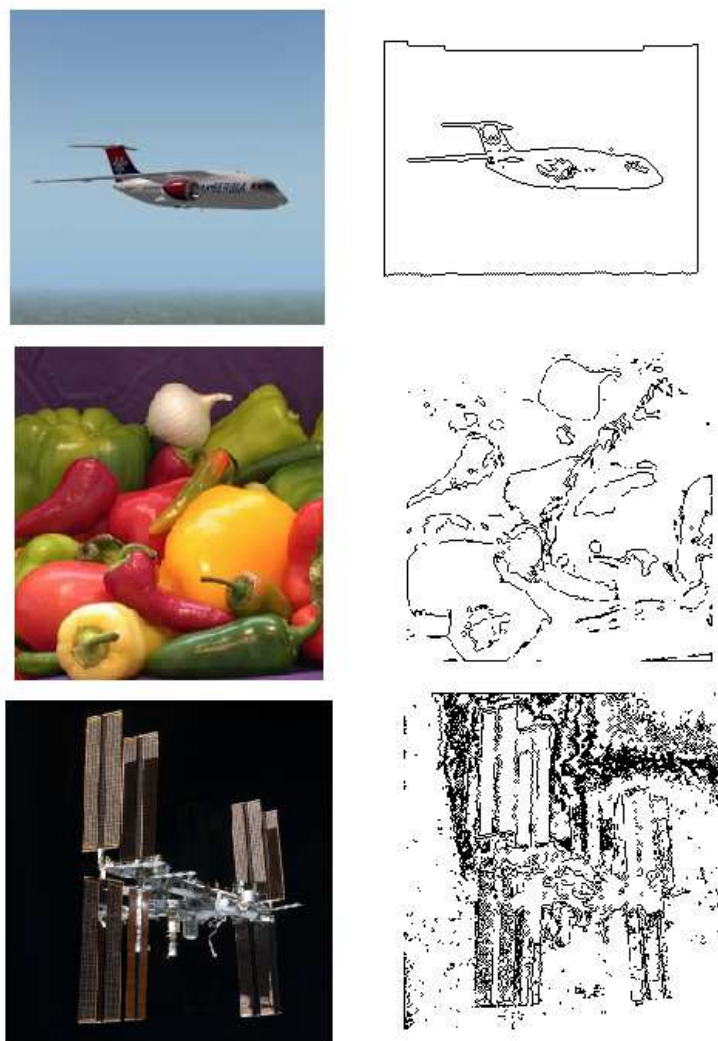


Рис. 1. Примеры автоматической кластеризации изображений



Рис. 2. Пример работы алгоритма в интерактивном режиме на изображении «Самолёт»

ЛИТЕРАТУРА

1. Kuntimad G., Ranganath H.S. Perfect image segmentation using pulse coupled neural networks // IEEE Transactions on Neural Networks. 1999. V. 10, No. 3. P. 591–597.
2. Yide M., Qing L. Automated image segmentation using improved PCNN model based on cross-entropy // International Symposium on Intelligent Multimedia, Video and Speech Processing. 2004. P. 20–22.
3. Yide M., Chunliang Q. Study of Automated PCNN System Based on Genetic Algorithm // Journal of system simulation. 2006. V. 18, No. 3. P. 722–724.
4. Xiaodon G., Shide G., Daohen Y. A new approach for image segmentation based on unit-linking PCNN // Proceeding of the first International Conference on Machine Learning and Cybernetics. 2002.
5. Stewart R.D.F., Oppor M. Region growing with pulse-coupled neural networks: An alternative to seeded region growing // IEEE Transactions on Neural Networks. 2002. V. 13, No. 6. P. 1557–1562.
6. Qing Z., Guanhui Y., Tingling G., Hong Z., Junxiao L. Fabric Defect Segmentation Based on Region Growing PCNN Model // Computer application and software. 2011. V. 28, No. 11. P. 171–175.
7. Ma H.-R., Cheng X.-W. Automatic Image Segmentation with PCNN Algorithm Based on Grayscale Correlation // International Journal of Signal Processing, Image Processing and Pattern Recognition. 2014. V. 7, No. 5. P. 249–258.
8. Привалов М.В., Скобцов Ю.А., Кудряшов А.Г. Применение нейросетевых методов классификации для сегментации компьютерных томограмм // ВЕСТНИК ХНТУ. 2010. № 2(38). С. 103–109.
9. Chandhok C. A Novel Approach to Image Segmentation using Artificial Neural Networks and K-Means Clustering // International Journal of Engineering Research and Applications (IJERA). 2012. V. 2, No. 3. P. 274–279.
10. Xu B., Lin S. Automatic Color Identification in Printed Fabric Images by a Fuzzy Neural Network // Computer Journal of AATICC Review. 2002. V. 2, No. 9. P. 42–45.
11. Yao K., Mignotte M., Collet C., Galerne P., Burel G. Unsupervised Segmentation Using a Self Organizing Map and a Noise Model Estimation in Sonar Imagery // Computer Journal of Pattern Recognition Letters. 2000. V. 33, No. 9. P. 1575–1584.
12. Aria E., Saradjian M., Amini J., Lucas C. Generalized Cooccurrence Matrix to Classify IRS-1d Images Using Neural Networks // Proceedings of ISPRS Congress. 2004. P. 117–123..
13. Awad M., Chehdi K., Nasri A. Multi Component Image Segmentation Using Genetic Algorithm and Artificial Neural Network // Computer Journal of Geosciences and Remote Sensing Letters. 2007. V. 4, No. 4. P. 571–575.
14. Zhou Z., Wei S., Zhang X., Zhao X. Remote Sensing Image Segmentation Based on Self Organizing Map at Multiple Scale // Proceedings of SPIE Geoinformatics: Remotely Sensed Data and Information. 2007. P. 122–126.
15. Kurnaz M., Dokur Z., Olmez T. Segmentation of Remote Sensing Images by Incremental Neural Network // Computer Journal of Pattern Recognition Letters. 2005. V. 26, No. 8. P. 1096–1104.
16. Белим С.В., Кутлуниин П.Е. Выделение контуров на изображениях с помощью алгоритма кластеризации // Компьютерная оптика. 2015. Т. 39, № 1. С. 119–124.

17. Белим С.В., Ларионов С.Б. Алгоритм сегментации изображений, основанный на поиске сообществ на графах // Компьютерная оптика. 2016. Т. 40, № 6. С. 904–910.

ALGORITHM OF A LEARNING SET FORMATION FOR IMAGE SEGMENTATION ARTIFICIAL NEURAL NETWORK

S.V. Belim

Dr.Sc. (Phys.-Math.), Professor, e-mail: sbelim@mail.ru

S.B. Larionov

Postgraduate Student, e-mail: me@stas-larionov.ru

Dostoevsky Omsk State University, Omsk, Russia

Abstract. An algorithm for forming a learning set for an artificial neural network in the problem of image segmentation is proposed in the article. A distinctive feature is the use of only one segmented image. Segmentation is performed using a three-layer perceptron. The method of growing areas is used. The neural network is used to decide whether the pixel belongs to the segment being formed. Impulse noise is used to form the learning set. Pixels modified by noise do not belong to any segment. A computer experiment was conducted in an automatic and interactive mode.

Keywords: image segmentation, neural network, impulse noise.

Дата поступления в редакцию: 30.03.2018

СРАВНЕНИЕ ИЗОБРАЖЕНИЙ НА ОСНОВЕ ПОСТРОЕНИЯ УРАВНЕНИЙ ГАМИЛЬТОНА

С.В. Лейхтер¹

аспирант, e-mail: leykhter@mail.ru

С.Н. Чуканов²

профессор, д.т.н., ведущий научный сотрудник, e-mail: ch_sn@mail.ru

¹ФГБОУ ВО Сибирский государственный автомобильно-дорожный университет
(СибАДИ), Омск, Россия

²ФГБУН Институт математики им. С.Л. Соболева СО РАН, Россия

Аннотация. В работе рассмотрена задача сравнения изображений кривых, представленных набором точек-ориентиров. Для анализа деформаций изображения в работе рассматривается группа диффеоморфизмов. Задача решается на основе метода построения функционала, характеризующего эволюцию диффеоморфного преобразования изображения от начального до требуемого состояния, и «штрафа» за отклонение траекторий точек изображения от необходимых. Приводится постановка задачи, основанная на построении и решении уравнений Гамильтона для группы диффеоморфизмов точечных ориентиров изображения. Разработан алгоритм решения уравнений Гамильтона диффеоморфного преобразования на основе метода стохастического градиентного спуска.

Ключевые слова: распознавание образов, стохастический градиентный спуск, уравнение Гамильтона, диффеоморфные преобразования.

Введение

Существуют различные методы решения задачи сравнения форм исходного изображения и терминального (целевого) изображения [14]. Для анализа деформаций изображения от исходного к целевому в работе рассматривается группа диффеоморфизмов, а деформация рассматривается как геодезическая (кратчайший путь) между исходным и целевым изображениями.

Дифференцируемое отображение $f : \mathcal{M} \rightarrow \mathcal{N}$ называется диффеоморфизмом для многообразий $\mathcal{M}, \mathcal{N}$, если оно является биективным и обратное отображение $f^{-1} : \mathcal{N} \rightarrow \mathcal{M}$ также дифференцируемое. В этом случае многообразия $\mathcal{M}, \mathcal{N}$ являются диффеоморфными ($\mathcal{M} \simeq \mathcal{N}$). Группа диффеоморфизмов дифференцируемого многообразия $\mathcal{M}$ — это группа всех диффеоморфизмов $\mathcal{M} \rightarrow \mathcal{M}$. В работах [11, 15, 16] используются уравнения В.И. Арнольда [3] для построения уравнений временной эволюции элементов групп диффеоморфизмов.

Задача сравнения исходного и целевого изображений решается на основе метода построения минимизируемого функционала [4, 5, 8, 11, 15, 16, 18, 20], ха-

рактеризующего эволюцию диффеоморфного преобразования изображения от начального до терминального, и «штрафа» за отклонение траектории изображения от требуемой траектории.

Для решения задачи сравнения применяется метод частиц-ориентиров (наиболее важных точек на изображении) [13]. В работе приводится постановка задачи, основанная на построении и решении уравнений Гамильтона для группы диффеоморфизмов частиц — точечных ориентиров изображения. Применение уравнений Гамильтона для группы диффеоморфизмов частиц позволяет не применять уравнения В.И. Арнольда для построения уравнений временной эволюции элементов групп диффеоморфизмов. Разработан алгоритм решения уравнений Гамильтона диффеоморфного преобразования на основе метода стохастического градиентного спуска.

Формирование диффеоморфизма изображений на основе всех точек (пикселей) изображения потребует большого объёма вычислительных операций. Использование пространств с метрикой С.Л. Соболева позволяет построить оптимальное векторное поле скоростей всех точек изображения по выбранному подмножеству точечных ориентиров, что значительно снижает требования к объёму вычислительных операций. Сформированное векторное поле в каждой точке изображения определяет диффеоморфное отображение исходного изображения как совокупности точечных ориентиров в терминальное (целевое) изображение.

1. Гамильтонова механика точечных ориентиров изображения

Диффеоморфизм изображений можно представить как эволюцию точечных ориентиров (наиболее важных точек) изображения на основе гамильтоновой механики. Рассмотрим параметризацию формы изображения с помощью частиц — точечных ориентиров. Пусть $q_i(t)$ используется для обозначения вектора положения i -й частицы и $p_i(t)$ соответствующего вектора импульса в момент времени t . Если формально принять, что импульсы и скорости частиц связаны соотношением $p_i = \mathcal{L} \cdot v_i$, где $\mathcal{L}$ — линейный обратимый оператор (в механике соответствует инертной массе), то обратное соотношение $v_i = \mathcal{L}^{-1} \cdot p_i = \mathcal{K} p_i$. Скалярное произведение векторов p_i и q_i , соответствующее метрике С.Л. Соболева [19], можно представить в виде $\langle p_i, v_i \rangle_{L^2} = \langle \mathcal{L} v_i, v_i \rangle_{L^2}$. Для оператора $\mathcal{L} = \text{id} - \alpha \nabla^2$ в пространстве $\Omega = \mathbb{R}^2$ обратным оператором формально является оператор $\mathcal{K} = \mathcal{L}^{-1}$, который можно аппроксимировать скалярной функцией Гаусса [17]

$$K(q_i - q_j) = \beta e^{-\alpha^{-2}(q_i - q_j)^T(q_i - q_j)}. \quad (1)$$

Тогда поля импульсов и скоростей в пространстве с метрикой С.Л. Соболева могут быть построены из соотношений

$$\begin{aligned} p_t(q) &= \sum_i \delta(q - q_i(t)) \cdot p_i(t); \\ v_t(q) &= \sum_i K(q - q_i(t)) \cdot p_i(t). \end{aligned} \quad (2)$$

Построим минимизируемый функционал, который соответствует деформации изображения, представленного совокупностью точечных ориентиров, в виде

$$J_0 = \frac{1}{2} \int_0^1 \left\{ \sum_{i,j} p_i^T K(q_i - q_j) p_j \right\} dt. \quad (3)$$

Задачу минимизации J_0 можно представить как задачу оптимального управления, связанную с гамильтонианом

$$H_0(q, p) = \frac{1}{2} \sum_{i,j} p_i^T K(q_i - q_j) p_j. \quad (4)$$

Временная эволюция системы определяется уравнениями Гамильтона [10, 11]

$$\begin{aligned} \dot{p}_i &= -\frac{\partial H_0}{\partial q_i} = -\sum_j p_j \nabla_1 K(q_i - q_j) p_j; \\ \dot{q}_i &= \frac{\partial H_0}{\partial p_i} = \sum_j K(q_i - q_j) p_j, \end{aligned} \quad (5)$$

где $\nabla_1 K(q_i - q_j)$ представляет собой градиент функции $(q_i, q_j) \rightarrow K(q_i - q_j)$ по отношению к первому операнду q_i . Используя для $K(q_i - q_j)$ скалярную функцию Гаусса (1), получим

$$\nabla_1 K(q_i - q_j) = -2\alpha^{-2}\beta(q_i - q_j) e^{-\alpha^{-2}(q_i - q_j)^T(q_i - q_j)}. \quad (6)$$

Если минимизируемый функционал J_0 принять в виде (3), а гамильтониан системы в форме

$$H(q, p) = H_0(q, p) + \sigma^{-2} \sum_i (\dot{q}_i - v_i(q))^2 = H_0(q, p) + \sigma^2 \sum_i p_i^2, \quad (7)$$

то уравнения Гамильтона примут вид

$$\begin{aligned} \dot{p}_i &= -\frac{\partial H}{\partial q_i} = -\sum_j p_j^T \nabla_1 K(q_i - q_j) p_j; \\ \dot{q}_i &= \frac{\partial H}{\partial p_i} = \sum_j K(q_i - q_j) p_j + \sigma^2 p_i. \end{aligned} \quad (8)$$

В функции Гамильтона (7) вводится штраф $\sigma^{-2} \sum_i (\dot{q}_i - v_i(q))^2$ при наличии рассогласования $(\dot{q}_i - v_i(q))$.

2. Определение начальных значений вектора импульса

Для системы дифференциальных уравнений $\dot{x}(t) = u(t, x(t))$, где $u(t, x)$ — функция непрерывная по t и равномерно непрерывная по Липшицу по x с условием $x(t_0) = x|_{t=t_0} = x_0$, существует непрерывно дифференцируемое решение $x(t) = \psi(t, t_0, x_0)$ на интервале $t \in [t_0 - \varepsilon, t_0 + \varepsilon]$ для некоторого $\varepsilon > 0$. Правые части дифференциальных уравнений (8) являются непрерывно дифференцируемыми по аргументам $(q_k, p_k); k = 1, \dots, N$, поэтому решения (8), рассматриваемые как функции $(q_k, p_k)(t) = \psi(t, q_k|_{t=t_0}, p_k|_{t=t_0})$, будут непрерывно дифференцируемыми по аргументам $(t_0, q_k|_{t=t_0}, p_k|_{t=t_0}); k = 1, \dots, N$.

Рассмотрим метод обучения в задаче нахождения диффеоморфизма $q_k(0) \rightarrow q_k(1)$ при минимизации функционала

$$J^q = \sum_{i=1}^N [q_i(t) - q_i(1)]^2 \rightarrow \min, \quad (9)$$

где $q(target) = (q_1(t), \dots, q_N(t))$ — целевое множество точек. Так как для системы дифференциальных уравнений (8) для вектора q заданы начальные условия $q(0)$ при $t = 0$, а для вектора p — терминальные $p(1) = (0, \dots, 0)$ при $t = 1$, то получаем двухточечную краевую задачу. Сведение этой задачи к задаче Коши может быть проведено различными способами: стрельбы, квазилинеаризации, инвариантного погружения и т. д. Требуемый вектор начальных значений $p(0) = (p_1(0), \dots, p_N(0))$ может быть найден методом градиентного спуска с градиентом функционала по вектору $p(0)$: $\nabla_{p(0)} J^q$.

Рассмотрим метод стохастического градиентного спуска (SGD Stochastic Gradient Descent) [6, 12] для решения двухточечной краевой задачи. Обозначим данные обучения как D , а функционал стоимости, определяемый по этим данным $J_D^q(p(0))$. Пусть T подмножество примеров в D , а $J_T^q(p(0))$ функционал, вычисленный по множеству примеров T . Пусть $\nabla_{p(0)} J_D^q$ и $\nabla_{p(0)} J_T^q$ обозначают градиенты функционалов, вычисленные по наборам данных обучения D и T . В традиционном подходе градиентного спуска используется весь набор данных D на этапе обновления вектора $p(0)$

$$p^{k+1}(0) = p^k(0) - \mu \cdot \nabla_{p(0)} J_D^q, \quad (10)$$

где k — номер итерации, μ — шаг градиентного спуска. В случае стохастического градиентного спуска на этапе обновления используется только один пример из D

$$p^{k+1}(0) = p^k(0) - \mu \cdot \nabla_{p(0)} J^q, \quad (11)$$

причём на каждой итерации используются различные примеры, выбранные случайным образом. В случае стохастического мини-пакета на этапе обновления используется подмножество примеров из D . При наличии нескольких локальных минимумов минимизируемого функционала с разной глубиной стохастический градиентный спуск позволяет «прыгать» от одного локального минимума к другому, то есть находить глобальный минимум.

Пример 1. Рассмотрим пример деформации объекта (изображения цифры «ноль») на основе интегрирования дифференциальных уравнений (8) при диффеоморфизме цифр, которые представлены 16-ю точками, равномерно расположенными на кривой символов. Начальные значения компонент вектора определялись на основе метода SGD. Интегрирование уравнений проводилось при $\alpha = \beta = 1$, $\sigma^2 = 1$. На рисунках 1, 2, 3 представлены результаты диффеоморфизма цифры «ноль» в цифры «девять», «один» и «пять», получившиеся значения минимизируемого функционала J_0 равны 9,1; 8,85 и 10,51 соответственно.

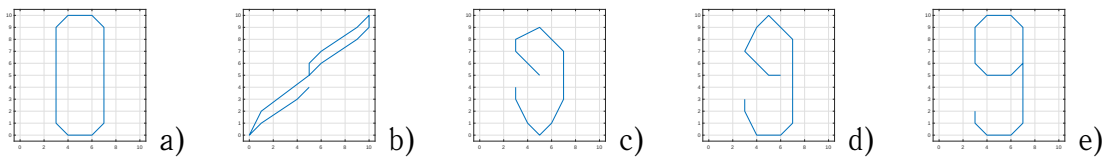


Рис. 1. Эволюция диффеоморфизма $0 \rightarrow 9$ в моменты времени: а) $t = 0,0$ (исходное изображение); б) $t = 0,25$; в) $t = 0,5$; д) $t = 0,75$; е) $t = 1,0$ (требуемый образ). $J_0 = 9,1$

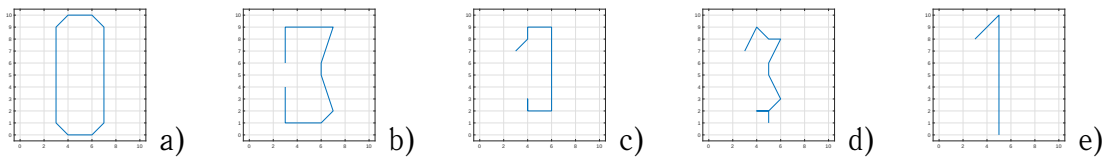


Рис. 2. Эволюция диффеоморфизма $0 \rightarrow 1$ в моменты времени: а) $t = 0,0$ (исходное изображение); б) $t = 0,25$; в) $t = 0,5$; д) $t = 0,75$; е) $t = 1,0$ (требуемый образ). $J_0 = 8,85$

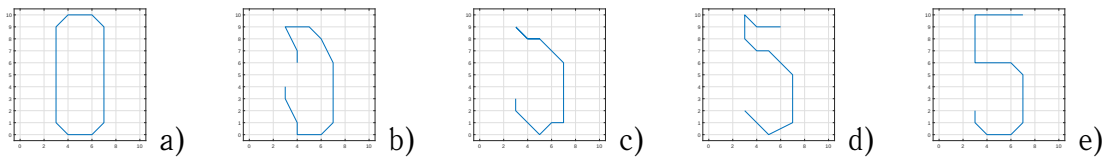


Рис. 3. Эволюция диффеоморфизма $0 \rightarrow 5$ в моменты времени: а) $t = 0,0$ (исходное изображение); б) $t = 0,25$; в) $t = 0,5$; д) $t = 0,75$; е) $t = 1,0$ (требуемый образ). $J_0 = 10,51$

Следует отметить, что исходные изображения имеют топологические характеристики отличные от терминальных. Например, число ручек изображения цифры «ноль» $g = 1$, а число ручек изображения цифры «пять» $g = 0$. $\square$

Существуют различные варианты искажения изображений: перенос (translation), вращение (rotation), масштабирование (scaling), перекокс (skew) и т. д. Алгоритм нормализации изображения преобразует исходное изображение в его нормальный (эталонный) вид. После нормализации может быть выполнено распознавание образа.

При известных эталонах изображений в базе данных изображений метод диффеоморфизма не требует предварительной нормализации [7, 19].

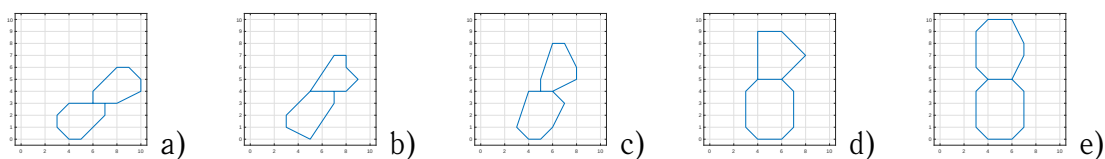


Рис. 4. Эволюция диффеоморфизма $s_{skew} \rightarrow 8$ в моменты времени: а) $t = 0,0$ (исходное изображение); б) $t = 0,25$; в) $t = 0,5$; д) $t = 0,75$; е) $t = 1,0$ (требуемый образ). $J_0 = 21,98$

Пример 2. На рисунке 4 приведён пример нормализации перекошенного (skewed) изображения цифры «восемь» к её нормальной форме.

Метод диффеоморфизма позволяет преобразовать перекошенное изображение к любому эталонному изображению, но при этом значение минимизируемого функционала будет больше, чем значение при деформации к нормальной форме цифры «восемь»; например, преобразование к нормальной форме цифры «четыре» требует минимизации функционала $J_0 = 21,98$.

3. Уравнения Гамильтона для отклонения траектории частиц от геодезических

Предположим, что для точечных ориентиров существует последовательность данных q^D ($q^D(t) = \{q_i^D(t)\}; i = 1, \dots, N$ — набор точечных ориентиров в момент времени t) и исходный шаблон $q_0 = q(0)$. Построим минимизируемый функционал, учитывающий отклонения траектории частиц (точечных ориентиров) от геодезических, в виде

$$J^d = \frac{1}{2} \int_0^1 \left\{ \sum_{i,j} p_i^T K(q_i - q_j) p_j \right\} dt + \frac{\lambda}{2} \int_0^1 \left\{ \sum_i (q_i^D - q_i)^2 \right\} dt. \quad (12)$$

Задачу минимизации функционала (12) можно представить как задачу оптимального управления частицами, связанную с гамильтонианом

$$H(p, q, t) = H_0(p, q) - \frac{\lambda}{2} \sum_i (q_i^D - q_i)^2, \quad (13)$$

где $H_0(p, q) = \frac{1}{2} \sum_{i,j} p_i^T K(q_i - q_j) p_j$, а член $\frac{\lambda}{2} \sum_i (q_i^D - q_i)^2$ штрафует за отклонение траектории частиц (точечных ориентиров). Соответствующими оптимальными траекториями являются решения уравнений Гамильтона

$$\begin{aligned} \dot{q}_i &= \frac{\partial H_0}{\partial p_i}; \\ \dot{p}_i &= -\frac{\partial H_0}{\partial q_i} + \lambda(q_i - q_i^D) \end{aligned} \quad (14)$$

или

$$\begin{aligned}\dot{q}_i &= \sum_j K(q_i - q_j) p_j; \\ \dot{p}_i &= - \sum_j p_i^T \nabla_1 K(q_i - q_j) p_j + \lambda (q_i - q_i^D).\end{aligned}\tag{15}$$

Оптимальное векторное поле $v(q(t))$ любой точки изображения q может быть восстановлено из соотношения

$$v(q(t)) = \sum_i K(q(t) - q_i(t)) p_i(t)$$

по выбранному подмножеству векторов точечных ориентиров q_i и соответствующих импульсов p_i , поэтому отсутствует необходимость решать уравнения (15) для всех точек изображения. Сформированное векторное поле в каждой точке изображения определяет диффеоморфное отображение исходного изображения как совокупность точечных ориентиров в терминальное изображение.

Заключение

В работе рассмотрена задача сравнения начального и терминального изображений, которая решается на основе построения минимизируемого функционала, характеризующего эволюцию диффеоморфного преобразования изображения от начального до терминального, и «штрафа» за отклонение траектории изображения от требуемой траектории. Разработан алгоритм решения уравнения диффеоморфного преобразования на основе метода стохастического градиентного спуска.

Приводится постановка задачи, основанная на решении уравнений Гамильтона для группы диффеоморфизмов. Вводится метод частиц-ориентиров (наиболее важных точек на изображении), который позволяет непосредственно решать задачу диффеоморфизма точечных ориентиров изображения.

Результаты работы для решения задачи детерминированной эволюции формы могут быть распространены для изучения стохастического возмущения гамильтоновых уравнений. Эволюцию точечных ориентиров можно рассматривать как систему частиц с внедрением случайной силы на каждую частицу, которая обеспечивает временное случайное возмущение относительно средней геодезической траектории [1, 2, 9].

Благодарности

Работа выполнена при поддержке комплексной программы фундаментальных научных исследований СО РАН I.5.1.7, проект 0314-2016-0020, и при финансовой поддержке Российского фонда фундаментальных исследований (гранты № 18-07-00526 и № 18-08-01284)

ЛИТЕРАТУРА

1. Arnaudon A., Holm D.D., Sommer S. Stochastic metamorphosis with template uncertainties. arXiv preprint arXiv:1711.07231. 2017. 16 p. URL: <https://arxiv.org/pdf/1711.07231.pdf> (дата обращения: 10.04.2018).
2. Arnaudon A., Holm D.D., Sommer S. A geometric framework for stochastic shape analysis. arXiv preprint arXiv:1703.09971. 2017. 47 p. URL: <https://arxiv.org/pdf/1703.09971.pdf> (дата обращения: 10.04.2018).
3. Arnold V. Sur la geometrie differentielle des groupes de Lie de dimension infinie et ses applications a l'hydrodynamique des fluides parfaits // Annales de l'institut Fourier. 1966. V. 16, No. 1. P. 319–361.
4. Ashburner J., Miller M.I. Diffeomorphic Image Registration // Brain Mapping: An Encyclopedic Reference. 2015. V. 1. P. 315–321.
5. Beg M., Miller M., Troune A., Younes L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms // Int. Journal of Computer Vision. 2005. V. 61, No. 2. P. 139–157.
6. Camassa R., Kuang D., Lee L. A Geodesic Landmark Shooting Algorithm for Template Matching and Its Applications // SIAM Journal on Imaging Sciences. 2017. V. 10, No. 1. P. 303–334.
7. Chukanov S.N. Definitions of invariants for n-dimensional traced vector fields of dynamic systems // Pattern Recognition and Image Analysis. 2009. V. 19, No. 2. P. 303–305.
8. Glaunes J., Qiu A., Miller M., Younes L. Large deformation diffeomorphic metric curve mapping // International journal of computer vision. 2008. V. 80, No. 3. P. 317–336.
9. Holm D.D. Stochastic metamorphosis in imaging science. arXiv preprint arXiv:1705.10149. 2017. 17 p. URL: <https://arxiv.org/pdf/1705.10149.pdf> (дата обращения: 10.04.2018).
10. Holm D.D., Schmah T., Stoica C. Geometric mechanics and symmetry: from finite to infinite dimensions. Oxford University Press, 2009. 537 p.
11. Holm D.D., Troune A., Younes L. The Euler-Poincare theory of metamorphosis // Quarterly of Applied Mathematics. 2009. V. 67, No. 4. P. 661–685.
12. Ketkar N. Deep Learning with Python: A Hands-on Introduction. Apress. 2017. 164 p.
13. Marsland S., McLachlan R.I. A Hamiltonian particle method for diffeomorphic image registration // Proceedings of Information Processing in Medical Images. Lect Notes Comput Sci. New York : Springer. 2006. P. 396–407.
14. Miller M., Younes L. Group actions, homeomorphisms, and matching: a general framework // Int. Journal of Computer Vision. 2001. V. 41, No. 1–2. P. 61–84.
15. Miller M., Troune A., Younes L. On metrics and the Euler-Lagrange equations of computational anatomy // Annual Reviews in Biomedical Eng. 2002. No. 4. P. 375–405.
16. Troune A., Younes L. Metamorphoses through lie group action // Foundations of Computational Mathematics. 2005. V. 5, No. 2. P. 173–198.
17. Younes L., Arrate F., Miller M.I. Evolutions equations in computational anatomy // NeuroImage. 2009. V. 45, No. 1. P. S40–S50.
18. Лейхтер С.В., Чуканов С.Н. Сравнение изображений на основе их диффеоморфного преобразования // Компьютерная оптика. 2018. Т. 42, №. 1. С. 96–104.

19. Чуканов С.Н., Ульянов Д.В. Формирование инвариантов при визуализации векторных полей на основе построения оператора гомотопии // Компьютерная оптика. 2012. Т. 36, №. 4. С. 622–626.
20. Лейхтер С.В., Чуканов С.Н. Построение метаморфизмов растровых изображений на основе решения уравнений Эйлера-Пуанкаре // Математические структуры и моделирование. 2017. Т. 43, №. 3. С. 86–95.

THE MATCHING OF IMAGES BASED ON THE CONSTRUCTION OF THE HAMILTON EQUATIONS

S.V. Lejhter¹

Graduate Student, e-mail: leykhter@mail.ru

S.N. Chukanov²

Dr.Sc. (Eng.), Professor, Leading Scientific Employee, e-mail: ch.sn@mail.ru

¹State Automobile and Highway University, Omsk, Russia

²Sobolev Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences, Russia

Abstract. The problem of comparing the template and the terminal images is considered. To analyze deformations of a image, the group of diffeomorphisms is considered. The problem is solved on the basis of the method for constructing a minimized functional characterizing the evolution of the diffeomorphic image transformation from the template to the terminal image, and the penalty for deviating the image path from the required trajectory. The formulation of the problem based on the construction and solution of Hamilton's equations for the group of diffeomorphisms of particles — points of landmark of the image is given. An algorithm for solving Hamilton's equations for a diffeomorphic transformation is developed based on the stochastic gradient descent method.

Keywords: pattern recognition, stochastic gradient descent, Hamilton equation, diffeomorphic transformation.

Дата поступления в редакцию: 11.05.2018

ПОДХОДЫ К РЕАЛИЗАЦИИ СЕТЕВОГО ПРОТОКОЛА ОБЕСПЕЧЕНИЯ ГАРАНТИРОВАННОЙ ДОСТАВКИ ПРИ МУЛЬТИМАРШРУТНОЙ ПЕРЕДАЧЕ ДАННЫХ

С.В. Гусс

старший преподаватель, e-mail: infoguss@gmail.com

Д.Н. Лавров

к.т.н., доцент, e-mail: lavrov@omsu.ru

Омский государственный университет им. Ф.М. Достоевского, Омск, Россия

Аннотация. В работе рассматривается проблема повышения надёжности передачи данных в сети с мерцающими узлами. Даются некоторые замечания по разработке алгоритмов разделения потока данных и восстановления потерянных данных. Предлагаются общие соображения, связанные с реализацией протокола многопутевой маршрутизации в самоорганизующихся динамических сетях.

Ключевые слова: маршрутизация, разделение секрета, восстановление данных, пропускная способность, сетевая отказоустойчивость.

Введение

В данной работе под узлом прежде всего понимается промежуточное устройство, способное передавать трафик. Мерцающий указывает на нестабильность работы, либо непостоянство присутствия в сети, возможно, в результате перемещения в пространстве. Предполагается, что надёжность передачи можно повысить, взяв за основу методы разделения секрета и коды, корректирующие ошибки.

Гарантированная доставка всегда была актуальной проблемой для передачи важных данных. Этот вопрос решается на нескольких уровнях модели OSI.

Традиционно можно начать с транспортного уровня. Здесь гарантированную доставку обеспечивает протокол ТСП, используется механизм подтверждения полученных сегментов данных. Протокол не учитывает наличия нескольких маршрутов до сети места назначения и не использует возможность более быстрой доставки за счёт разделения сообщения. Последнюю можно было бы использовать для введения избыточности в целях повышения надёжности. В случае с ТСП избыточность может привести к нестабильности работы протокола и некорректной его работе.

На сетевом уровне возможна балансировка нагрузки, которая осуществляется в большинстве случаев для равновесных маршрутов. Осуществление гарантированной доставки возлагается на протоколы вышележащих уровней.

На канальном уровне в пределах локальной сети решение возможно за счёт резервирования. Помимо надёжности резервирование в некоторых случаях способствует ускорению передачи данных, например, за счёт агрегирования каналов (технология EtherChannel), в связи с чем и создаётся возможность балансировки. Протокол STP тоже создаёт резервные пути при построении связующего дерева, но балансировка нагрузки в итоговом дереве отсутствует.

Существует ещё возможность резервирования и балансировки шлюзов с помощью протоколов HSRP и GLBP.

Рассмотрим проблему гарантированной доставки в таком варианте. Передача данных осуществляется через агрессивную среду или среду с отсутствием доверия к отдельным узлам. В такой среде может присутствовать злоумышленник, контролирующий только часть узлов, через которые могут проходить данные. Злоумышленник может быть пассивным и пытаться только просмотреть данные, либо же активным: может пытаться повредить данные, изменить и, в конце концов, не дать доставить данные до места назначения. Если первые два действия ещё могут быть защищены стандартными криптографическими средствами и технологиями типа VPN, то от последнего из перечисленных действий шифрование и алгоритмы контроля целостности не защитят — сообщение просто не сможет быть доставлено.

В настоящее время существуют подозрения, что часть выходных узлов той же TOR-сети контролируются недружественными иностранными организациями. Тогда появляется задача написать TOR-клиент, который будет с помощью алгоритмов разделения секрета разделять потоки между несколькими выходными узлами TOR-сети и таким образом гарантировать доставку до места назначения, а сами алгоритмы разделения секрета помогут восстановить уничтоженные или потерянные данные. Описанный подход может быть применён для произвольной группы прокси-серверов, не обязательно TOR-сети.

1. Сети с мерцающими узлами

Концепция мерцающих узлов более применима для так называемых крупных подвижных объектов — кораблей, самолётов, поездов [1]. А сами узлы обслуживаются системами спутниковой связи. Обсуждение проблемы ведётся с точки зрения физического доступа и связи. Существуют определённые наработки в виде топологических моделей систем связи, которые также применимы к самоорганизующимся компьютерным сетям. Эти модели оказывают влияние на логику работы алгоритмов маршрутизации. Протоколы маршрутизации в таких системах могут учитывать повторения во времени топологических изменений. Главная особенность алгоритмов — невозможность гарантировать передачу информации по заданному маршруту и графику движения, плюс ко всему требуются постоянные корректировки таблиц маршрутизации, а также учёт статистических данных. В логике маршрутизации можно учитывать повторения, длительность, стабильность состояний расположения сетевых устройств и разнородность их скоростей и производительности. Может производиться «формирование базы данных характеристик движения объектов и расчёт времени

открытия и закрытия каналов между парами устройств» [1].

2. Методы разделения секрета

Предполагается разделение потока информации на несколько частей (потоков) и одновременная их передача по разным каналам связи. Это позволит наиболее полно задействовать ресурсы сети и снизить риск перехвата и восстановления отправленного сообщения [2]. Интересным решением видится обеспечение взаимозависимости между потоками. В идеале это должно усложнить восстановление сообщения в случае перехвата. Отмечается, что при передаче текстовой информации в некоторых случаях возможно задействовать особенности естественного языка, связанные с частотой появления символов и существованием зависимости между близко стоящими символами. Если принять во внимание две эти особенности, можно улучшить надёжность разделения потоков с точки зрения безопасности. В работе [3] предлагается «предварительно использовать один из оптимальных алгоритмов кодирования, позволяющий выравнивать статистические характеристики текста».

3. Коды, корректирующие ошибки

Главной особенностью алгоритмов, способных обнаруживать и исправлять ошибки, возникающие в процессе передачи, является добавление избыточной информации. Всё это можно использовать не только в целях исправления ошибок и контроля целостности данных, но и для создания неблагоприятных условий для злоумышленника, пытающегося восстановить или внести изменения в передаваемую информацию. Сегодня всё чаще говорят о «внедрении интегрированных механизмов обеспечения безопасности и достоверности в протоколах обмена данными компьютерных сетей» [4].

4. Алгоритмы разделения и восстановления данных

В схемах разделения секрета при реализации защищённых распределённых вычислений традиционно присутствуют [5] постановщик задачи (тот, кто распределяет данные) и получатель результата. Первый отвечает за разделение данных и за соответствующие вопросы (избыточность, форматирование, шифрование), а второй за восстановление данных. Традиционно распределённость вычислений предполагает наличие нескольких интеллектуальных узлов, обрабатывающих информацию по определённым алгоритмам.

5. Реализация протокола многопутевой маршрутизации в самоорганизующихся динамических сетях

Алгоритм построения маршрутов в компьютерной сети зависит от выбора критерия оптимизации [6]. Учитывается изменение топологии (особенно в самоорганизующихся сетях) и нагрузка на сеть. Для стабильных сетей, с малой

или нулевой степенью изменения, маршруты строятся заранее, на что требуется определённое время для вычислений. Такой подход не годится, когда нужно по возможности быстро реагировать на изменения в топологии. Поскольку заранее не известно количество узлов сети и качество каналов связи между ними, задача построения маршрутов становится ещё сложнее.

Имеет смысл накапливать информацию в процессе работы протокола маршрутизации в конкретных условиях. Это необходимо для выявления определённой модели поведения и составления набора профилей. В процессе работы протокола такой профиль выбирается автоматически на основе анализа ситуации, определении паттерна действий и происходит настройка процесса маршрутизации согласно выбранному алгоритму, пригодному в данном конкретном случае [7]. Таким образом, предлагается подход, в котором критерий оптимизации определяется в процессе использования сетевых ресурсов. На основе критерия оптимизации выбирается наиболее приемлемый профиль, влияющий на процесс построения маршрута путём запуска того или иного алгоритма в зависимости от количества узлов в сети и её топологии на данный момент.

Реализация (программная или аппаратная) любого сетевого протокола возможна только после того, как составлено описание алгоритма его работы: формат и последовательность сообщений, временные характеристики, условия выбора, очерёдность принятия решений. Описание протокола — общая схема, по которой могут взаимодействовать реализации различных производителей. Если протокол востребован и успешен, то возможно множество его реализаций, отличающихся стабильностью работы, удобством настройки, полнотой реализации для той или иной системы (например, операционной). Общедоступное, полное описание протокола создаётся тогда, когда он прошёл ряд испытаний и показал свою пригодность на практике. Цель исследований авторов статьи — получить в конечном итоге реализацию протокола динамической многопутевой маршрутизации в сети с непостоянной, быстроменяющейся топологией.

В многопутевой маршрутизации (multipath routing) для передачи трафика могут использоваться несколько маршрутов, пролегающих через разные каналы связи. Большая нагрузка на быстрый канал связи может сделать весь маршрут не привлекательным с точки зрения надёжности и отказоустойчивости в отсутствие правильной политики выставления приоритетов на тот или иной тип трафика. В тоже время, не факт, что канал будет нагружен постоянно. Таким образом, алгоритм построения маршрута должен основываться не только на качестве каналов связи, составляющих маршрут, но и на статистике использования каналов при построении резервных и, в идеальном случае, параллельных путей для балансировки нагрузки и ускорения передачи данных, когда есть возможность по максимуму использовать ресурсы ненагруженной сети. Вообще говоря, в этом и заключается основная задача многопутевой маршрутизации — оптимизация использования ресурсов сети, а это приводит к избавлению от перегрузок и повышению стабильности сети.

В качестве инструмента имитационного моделирования для проверки работоспособности протокола предполагается мультиагентное моделирование (agent-based modeling). В качестве объекта могут быть выбраны беспровод-

ные компьютерные, самоорганизующиеся сети (mesh, ad-hoc), тем не менее, протокол не ограничивается исключительно беспроводными сетями. Последние выбраны в связи с их динамическим характером изменения.

Для проведения исследований принято решение разработать симулятор [8], возможно, в качестве надстройки (расширения) над существующей системой имитационного моделирования. Такая система должна обладать возможностью описания агентов, проведения экспериментов и оценки функционирования. Описание среды и объектов должно проходить на трёх уровнях: микро (низкий уровень абстракции, для учёта деталей), мезо (средний), макро (высокий). В свою очередь сам процесс моделирования должен проходить на трёх уровнях. На оперативном уровне задаются параметры (моделирование физической среды): размеры, расстояния, время, скорость. На тактическом уровне составляются сценарии: возможное изменение параметров и характеристик объектов, передвижение, маршрутизация трафика (расписание, задержки). На стратегическом уровне, где происходят взаимодействие, препятствия, влияния, внешние воздействия, обратные связи, тенденции — производится вмешательство в моделируемую среду, введение непредвиденных обстоятельств. Моделирование проходит в определённой последовательности: задание характеристик, сценария (миссии) передвижения, потоков трафика (кто, кому); выбор карты, местности, локации с объектами (статическими, динамическими); выбор алгоритмов поведения агентов в различных условиях и вмешательство в процесс по ходу действия.

Аналитические способности симуляции: идентификация и прогнозирование состояния системы, рекомендации (стратегические, тактические, оперативные).

Таким образом, реализация сетевого протокола маршрутизации предполагает разработку (или расширение) программной системы имитационного моделирования. Ожидаемые возможности такого средства: выбор или создание агентов (устройства, препятствия), задание параметров и проверка соответствия стандартам и рекомендациям, возможность записи алгоритмов и сценариев на специализированном упрощённом языке (сценариев или графическом), визуализация перемещения участников сетевого взаимодействия.

Стоит сделать ряд замечаний по поводу такого подхода к реализации протокола. Поскольку в ходе моделирования необходимо задать (выбрать или описать) формулу (статистику), необходимо разбираться в моделируемой предметной области, в которой происходит сетевое взаимодействие, для некоторых областей вероятен высокий уровень вхождения и сложность использования средства моделирования без консультаций аналитика или эксперта из предметной области. Поскольку обстоятельства в моделируемых ситуациях возникают из контекста, имитационная система должна быть достаточно интеллектуальной, чтобы соответствовать действительности и вполне вероятно, что какие-то аспекты реальности будут оставлены без внимания. В любом случае это только начальная стадия разработки, где моделирование происходит в упрощённой реальности взаимодействия агентов. И первые реализации протокола могут быть только прототипами, которые на начальном этапе своей работы должны собрать достаточно информации для построения профилей и продемонстрировать свою

способность в выборе оптимального алгоритма составления маршрутов. Всё это послужит базой для получения численных результатов оценки производительности при различных схемах и сценариях сетевого взаимодействия.

Заключение

В работе была рассмотрена проблема реализации протокола гарантированной доставки данных в ненадёжной среде передачи с точки зрения прослушивания и непостоянства топологии. В планах стоят задачи реализации протокола многопутевой маршрутизации и разработка программного клиента, который будет с помощью алгоритмов разделения секрета разделять потоки между несколькими выходными узлами сети и таким образом гарантировать доставку до места назначения, а сами алгоритмы разделения секрета помогут восстановить уничтоженные или потерянные данные. Описанный подход может быть применён для произвольной группы прокси-серверов.

Благодарности

Работа выполнена в рамках гранта НИОКТР № 01-06/683.

ЛИТЕРАТУРА

1. Сорокин А.А., Дмитриев В.Н. Описание систем связи с динамической топологией сети при помощи модели мерцающего графа // Вестник АГТУ. Сер.: Управление, вычислительная техника и информатика. 2009. № 2. С.134–139.
2. Дулькейт В.И., Лавров Д.Н., Михайлов П.И., Свенч А.А. Анализ надёжности алгоритма разделения секрета в сетевых потоках // Математические структуры и моделирование. 2003. Вып. 12. С. 146–154.
3. Лавров Д.Н. Схема разделения секрета для потоков данных маршрутизируемой сети // Математические структуры и моделирование. 2002. Вып. 10. С. 192–197.
4. Сумцов Д.В., Евсеев С.П., Томашевский Б.П., Король О.Г. Эффективность обмена данными в компьютерной сети при различных способах управления обменом // Збірник наукових праць ДонІЗТ. 2009. № 17. С 33–45.
5. Файзуллин Р.Т., Щерба Е.В., Волков Д.А. Схема реализации параллельных вычислений как инструмент защиты обрабатываемых данных // Доклады ТУСУРа. 2014. № 2(32). С. 195–201.
6. Шувалов В.П., Варакина И.Ю. Классификация методов многопутевой маршрутизации // Т-СОММ: Телекоммуникации и транспорт. 2014. Т. 8, № 1. С. 29–32.
7. Гусс С.В. Самоорганизующиеся mesh-сети для частного использования // Математические структуры и моделирование. 2016. № 4(40). С. 102–115.
8. Гусс С.В. Имитационное моделирование беспроводных самоорганизующихся сетей с быстро меняющейся топологией // Математическое и компьютерное моделирование : сборник материалов IV Международной научной конференции (Омск, 11 ноября 2016 г.) / [отв. за вып. И.П. Бесценный]. Омск : Изд-во Ом. гос. ун-та, 2016. 111–112 с.

APPROACHES TO IMPLEMENTING A SECURE DELIVERY NETWORK PROTOCOL FOR MULTIPATH DATA TRANSFER

S.V. Guss

Senior Teacher, e-mail: infoguss@gmail.com

D.N. Lavrov

Ph.D.(Eng.), Associate Professor, e-mail: lavrov@omsu.ru

Dostoevsky Omsk State University, Omsk, Russia

Abstract. The paper considers the problem of increasing the reliability of data transmission in a network with flickering nodes. Some comments are given on the development of algorithms for dividing the data flow and recovering lost data. We propose general considerations related to the implementation of the protocol for multipath routing in self-organizing dynamic networks.

Keywords: routing, secret sharing, data recovery, bandwidth, network fault tolerance.

Дата поступления в редакцию: 16.05.2018

WHY DEEP LEARNING METHODS USE KL DIVERGENCE INSTEAD OF LEAST SQUARES: A POSSIBLE PEDAGOGICAL EXPLANATION

Olga Kosheleva

Ph.D. (Phys.-Math.), Associate Professor, e-mail: olgak@utep.edu

Vladik Kreinovich

Ph.D. (Phys.-Math.), Professor, e-mail: vladik@utep.edu

University of Texas at El Paso, El Paso, Texas 79968, USA

Abstract. In most applications of data processing, we select the parameters that minimize the mean square approximation error. The same Least Squares approach has been used in the traditional neural networks. However, for deep learning, it turns out that an alternative idea works better — namely, minimizing the Kullback-Leibler (KL) divergence. The use of KL divergence is justified if we predict probabilities, but the use of this divergence has been successful in other situations as well. In this paper, we provide a possible explanation for this empirical success. Namely, the Least Square approach is optimal when the approximation error is normally distributed — and can lead to wrong results when the actual distribution is different from normal. The need to have a robust criterion, i.e., a criterion that does not depend on the corresponding distribution, naturally leads to the KL divergence.

Keywords: Deep learning, Kullback-Leibler divergence.

1. Formulation of the Problem

Machine learning: reminder. The main problem of machine learning is:

- given input-output patterns $(x^{(k)}, y^{(k)})$,
- to come up with a function $f(x)$ for which $f(x^{(k)}) \approx y^{(k)}$.

This function can then be used to predict the output y for other inputs x .

In each model of machine learning:

- we have a function $f(x, c)$ depending on some parameters $c = (c_1, c_2, \dots)$, and
- we need to find the values of these parameters for which the resulting values $z^{(k)} \stackrel{\text{def}}{=} f(x^{(k)}, c)$ are approximately equal to the given value $y^{(k)}$:

$$z^{(k)} \approx y^{(k)}.$$

How to describe this approximate equality: traditional approach. Traditionally, in machine learning, the Least Squares approach was used to describe the desired approximate equality of:

- the result $z^{(k)}$ of applying the model $f(x, c)$ to the input $x^{(k)}$ and

• the given outputs $y^{(k)}$;
 see, e.g., [1]. Specifically, most traditional methods minimize the sum

$$\sum_{k=1}^K (z^{(k)} - y^{(k)})^2. \quad (1)$$

Deep learning techniques use KL divergence instead. In deep learning, it turned out that better results are obtained if, instead of the least squares technique (1), we use the Kullback-Leibler (KL) divergence; see, e.g., [2, 3]. Specifically, we re-scale the values $y^{(k)}$ and $z^{(k)}$ so that these values are always between 0 and 1, and then minimize the following objective function:

$$-\sum_{k=1}^K [y^{(k)} \cdot \log(z^{(k)}) + (1 - y^{(k)}) \cdot \log(1 - z^{(k)})]. \quad (2)$$

Why? At first glance, the least squares is a reasonable criterion, the one most frequently used in statistical data processing; see, e.g., [4]. So why is the alternative approach working better?

For the case when the predicted values $y^{(k)}$ are probabilities, an explanation is given in [2], Section 5.5. However, the criterion (2) is also successfully used in many applications in which the predicted values $y^{(k)}$ are not probabilities. How can we explain this success?

What we do in this paper. In this paper, we extend the existing probability-case explanation to the general case, thus providing a possible explanation of why KL divergence works well.

2. Our Explanation

Why least squares: reminder. In order to explain why KL divergence is efficient, let us first recall why the Least Squares method is often used.

Ideally, the deviations $z^{(k)} - y^{(k)}$ should be all 0s, but in reality, we can only attain an approximate equality. In different situations, we get different values of these deviations. It is therefore reasonable to view these deviations as random variables.

In practice, many random variables are normally distributed; see, e.g., [4]. It is therefore reasonable to assume that the deviations $z^{(k)} - y^{(k)}$ are normally distributed, with 0 means and some standard deviation σ . The corresponding probability density function is thus equal to

$$\rho(z^{(k)} - y^{(k)}) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(z^{(k)} - y^{(k)})^2}{2\sigma^2}\right). \quad (3)$$

Since we do not have any reason to believe that different deviations are positively or negatively correlated, it is reasonable to assume that different deviations are

independent. In this case, for each tuple c , the probability (density) is equal to the product of the corresponding probabilities (3), i.e., equal to

$$\rho(c) = \prod_{k=1}^K \left[\frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp \left(-\frac{(z^{(k)} - y^{(k)})^2}{2\sigma^2} \right) \right], \quad (4)$$

where $z^{(k)} = f(x^{(k)}, c)$. It is reasonable to select the tuple c which is the most probable, i.e., for which the expression (4) is the largest possible; this natural idea is known as the *Maximum Likelihood approach*.

Maximizing the expression (4) is equivalent to minimizing its negative logarithm

$$-\ln(\rho(c)) = \text{const} + \sum_{k=1}^K \frac{(z^{(k)} - y^{(k)})^2}{2\sigma^2},$$

and this minimization is equivalent to minimizing the Least Squares expression (1).

Need to go beyond the Least Squares. While many practical probability distributions are normal, there are also many cases when the probability distribution is different from normal; see, e.g., [4]. In such cases, the Least Squares method is not optimal — and it can be very far from optimal. For example, if we have a distribution with heavy tails, for which the probability of large deviations is high, the Least Squares method often leads to erroneous estimates.

This can be illustrated on the simple example when the model $f(x, c) = c_1$ is simply an unknown constant c_1 . In this case, if we minimize the sum $\sum_{k=1}^K (x^{(k)} - c_1)^2$ — by differentiating by c_1 and equating the derivative to 0 — we

get the estimate $\hat{c}_1 = \frac{1}{K} \cdot \sum_{k=1}^K y^{(k)}$. If the actual value of c_1 is, e.g., 10, and we get

$K = 100$ values close to 10, then the arithmetic average is indeed close to 10. But if one of the estimates is an outlier, e.g., $x^{(1)} = 10^6$, then the arithmetic average is close to 10,000 — way beyond the actual value 10.

To take this non-normality into account, we need to replace the Least Squares approach with a one which is *robust*, in the sense that it does not depend on the probability distribution of the divergence.

Main idea. In the computer, any real value is represented as 0s and 1s. To transform a real-valued signal into a sequence of 0s and 1s, measuring instruments use analog-to-digital converters. These converters are usually based on comparing the actual value with some threshold values. For example, if the actual value is between 0 and 1, then, by comparing this value with 0.5, we can tell whether the first bit in the binary expansion is 0 or 1:

- if the actual value is smaller than 0.5, then the first bit is 0, and
- if the actual value is larger than 0.5, then the first bit is 1.

By selecting a second threshold to be 0.25 or 0.75, we can determine the second bit, etc.

The thresholds are not necessarily binary-rational numbers: often, other thresholds are used, and then the resulting number is recovered from the results of the

corresponding comparisons.

We want to come up with a probabilistic interpretation; thus, it makes sense to select random thresholds. The simplest possible random number generator generates values uniformly distributed on the interval $[0, 1]$. Such random number generators are included in most programming languages.

Resulting setting. So, to describe each value $y^{(k)}$, let us run this simplest random number generator a large number of time N , and store N results of comparing $y^{(l)}$ with the corresponding random numbers r_i :

- we store 1 if $r_i \leq y^{(k)}$, and
- we store 0 if $r_i > y^{(k)}$.

For a random number uniformly distributed on the interval $[0, 1]$, the probability to be in each interval is equal to the width of this interval. In particular, the probability to be smaller than or equal to $y^{(k)}$ — i.e., the probability to be in the interval $[0, y^{(k)}]$ — is equal to $y^{(k)}$. Thus, for large N , we have:

- approximately $N \cdot y^{(k)}$ 1's and
- approximately $N \cdot (1 - y^{(k)})$ 0s.

For each of K patterns, we have N 0-1 records, so overall, we have a long sequence of $N \cdot K$ records corresponding to all K patterns. This sequence corresponds to the observations.

Derivation of KL divergence. We want to find the tuple c that best fits the above long sequence of observations.

For each tuple c and for each pattern k , the model $f(x, c)$ returns the value $z^{(k)} = f(x^{(k)}, c)$. Thus:

- the probability to get 1 when we compare this value with a random value r_i is equal to $z^{(k)}$, and
- the probability to get 0 is equal to the remaining probability $1 - z^{(k)}$.

So, for each pattern, we have:

- $N \cdot y^{(k)}$ observations with probability $z^{(k)}$, and
- $N \cdot (1 - y^{(k)})$ observations with probability $1 - z^{(k)}$.

Assuming — as before — that all observations are independent, we conclude that the probability of observing the given sequence of 0s and 1s is equal to the product of all these probabilities, i.e., to the value

$$(z^{(k)})^{N \cdot y^{(k)}} \cdot (1 - z^{(k)})^{N \cdot (1 - y^{(k)})}.$$

The overall probability can be obtained by multiplying probabilities corresponding to all K patterns. Thus, we select a tuple c for which the following expression is the largest possible:

$$p \stackrel{\text{def}}{=} \prod_{k=1}^K \left[(z^{(k)})^{N \cdot y^{(k)}} \cdot (1 - z^{(k)})^{N \cdot (1 - y^{(k)})} \right].$$

Maximizing this expression p is equivalent to minimizing its negative logarithm

$$-\ln(p) = - \sum_{k=1}^K \left[N \cdot y^{(k)} \cdot \ln(z^{(k)}) + N \cdot (1 - y^{(k)}) \cdot \ln(1 - z^{(k)}) \right].$$

This expression is N times larger than the KL divergence (3). Thus, minimizing this expression is indeed equivalent to minimizing the KL divergence.

So, we indeed get the desired explanation for minimizing the KL divergence.

Acknowledgments

This work was supported in part by the National Science Foundation grant HRD-1242122 (Cyber-ShARE Center of Excellence).

REFERENCES

1. Bishop C.M. Pattern Recognition and Machine Learning. New York : Springer Verlag, 2006.
2. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge : MIT Press, 2016.
3. Liu P., Choo K.-K.R., Wang L., Huang F. SVM or deep learning? A comparative study on remote sensing image classification // Soft Computing. 2017. V. 21. P. 7053–7065.
4. Sheskin D.J. Handbook of Parametric and Nonparametric Statistical Procedures. Boca Raton : Chapman and Hall/CRC, 2011.

ПОЧЕМУ МЕТОДЫ ГЛУБОКОГО ОБУЧЕНИЯ ИСПОЛЬЗУЮТ РАССТОЯНИЕ КУЛЬБАКА-ЛЕЙБЛЕРА ВМЕСТО МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ: ВОЗМОЖНОЕ ПЕДАГОГИЧЕСКОЕ ОБЪЯСНЕНИЕ

О. Кошелева

к.ф.-м.н., доцент, e-mail: olgak@utep.edu

В. Крейнович

к.ф.-м.н., профессор, e-mail: vladik@utep.edu

Техасский университет в Эль Пасо, США

Аннотация. В большинстве приложений обработки данных мы выбираем параметры, которые минимизируют среднеквадратичную ошибку приближения. Аналогичный метод наименьших квадратов использовался в традиционных нейронных сетях. Однако, оказалось, что для глубокого обучения лучше работает альтернативная идея, а именно: минимизация расстояния Кульбака-Лейблера. Использование расстояния Кульбака-Лейблера оправдано, если мы прогнозируем вероятности, но использование этого расстояния было успешным и в других ситуациях. В этой статье мы приводим возможное объяснение этого эмпирического успеха. А именно: метод наименьших квадратов оптимален, когда ошибка аппроксимации распределяется по нормальному закону, и может привести к неправильным результатам, когда фактическое распределение отличается от нормального. Необходимость иметь надежный критерий, т.е. критерий, который не зависит от соответствующего распределения, естественным образом приводит к расстоянию Кульбака-Лейблера

Ключевые слова: глубокое обучение, расстояние Кульбака-Лейблера.

Дата поступления в редакцию: 27.12.2017

HOW TO STORE TENSORS IN COMPUTER MEMORY: AN OBSERVATION

Martine Ceberio

Ph.D. (Phys.-Math.), Associate Professor, e-mail: mceberio@utep.edu

Vladik Kreinovich

Ph.D. (Phys.-Math.), Professor, e-mail: vladik@utep.edu

University of Texas at El Paso, El Paso, Texas 79968, USA

Abstract. In this paper, after explaining the need to use tensors in computing, we analyze the question of how to best store tensors in computer memory. Somewhat surprisingly, with respect to a natural optimality criterion, the standard way of storing tensors turns out to be one of the optimal ones.

Keywords: Tensors, computing, computer memory.

1. Why Tensors: A Reminder

Why tensors. One of the main problems of modern computing is that:

- we have to process large amounts of data;
- and therefore, long time is required to process this data.

A similar situation occurred in the 19 century physics:

- physicists had to process large amounts of data;
- and, because of the large amount of data, a long time is required to process this data.

We will recall that in the 19 century, the problem was solved by using tensors. It is therefore a natural idea to also use tensors to solve the problems with modern computing.

Tensors in physics: a brief reminder. Let us recall how tensors helped the 19 century physics; see, e.g., [6]. Physics starts with measuring and describing the values of different physical quantities. It goes on to equations which enable us to predict the values of these quantities.

A measuring instrument usually returns a single numerical value. For some physical quantities (like mass m), the single measured value is sufficient to describe the quantity. For other quantities, we need several values. For example, we need three components E_x , E_y , and E_z to describe the electric field at a given point. To describe the tension inside a solid body, we need even more values: we need 6 values $\sigma_{ij} = \sigma_{ji}$ corresponding to different values $1 \leq i, j \leq 3$: σ_{11} , σ_{22} , σ_{33} , σ_{12} , σ_{23} , and σ_{13} .

The problem was that in the 19 century, physicists used a separate equation for each component of the field. As a result, equations were cumbersome and difficult to solve.

The main idea of the tensor approach is to describe all the components of a physical field as a single mathematical object:

- a vector a_i ,
- or, more generally, a tensor $a_{ij}, a_{ijk}, \dots$

As a result, we got simplified equations — and faster computations.

It is worth mentioning that originally, mostly vectors (rank-1 tensors) were used. However, the 20 century physics has shown that higher-order matrices are also useful. For example:

- matrices (rank-2 tensors) are actively used in quantum physics,
- higher-order tensors such as the rank-4 curvature tensor R_{ijkl} are actively used in the General Relativity Theory.

From tensors in physics to computing with tensors. As we have mentioned earlier, 19 century physics encountered a problem of too much data. To solve this problem, tensors helped.

Modern computing suffers from a similar problem. A natural idea is that tensors can help. Two examples justify our optimism:

- modern algorithms for fast multiplication of large matrices; and
- quantum computing.

2. Modern algorithm for multiplying large matrices

In many data processing algorithms, we need to multiply large-size matrices:

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \dots & \dots & \dots \\ b_{n1} & \dots & b_{nn} \end{pmatrix} = \begin{pmatrix} c_{11} & \dots & c_{1n} \\ \dots & \dots & \dots \\ c_{n1} & \dots & c_{nn} \end{pmatrix}; \quad (1)$$

$$c_{ij} = a_{i1} \cdot b_{1j} + \dots + a_{ik} \cdot b_{kj} + \dots + a_{in} \cdot b_{nj}. \quad (2)$$

There exist many efficient algorithms for matrix multiplication.

The problem is that for large matrix size n , there is no space for both A and B in the fast (cache) memory. As a result, the existing algorithms require lots of time-consuming data transfers (“cache misses”) between different parts of the memory.

An efficient solution to this problem is to represent each matrix as a matrix of blocks; see, e.g., [2, 10]:

$$A = \begin{pmatrix} A_{11} & \dots & A_{1m} \\ \dots & \dots & \dots \\ A_{m1} & \dots & A_{mm} \end{pmatrix}, \quad (3)$$

then

$$C_{\alpha\beta} = A_{\alpha 1} \cdot B_{1\beta} + \dots + A_{\alpha\gamma} \cdot B_{\gamma\beta} + \dots + A_{\alpha m} \cdot B_{m\beta}. \quad (4)$$

Comment. For general arguments about the need to use non-trivial representations of 2-D (and multi-dimensional) objects in the computer memory, see, e.g., [21,22].

In the above idea,

- we start with a large matrix A of elements a_{ij} ;
- we represent it as a matrix consisting of block sub-matrices $A_{\alpha\beta}$.

This idea has a natural *tensor interpretation*:

- each element of the original matrix is now represented as
- an (x, y) -th element of a block $A_{\alpha\beta}$,
- i.e., as an element of a rank-4 tensor $(A_{\alpha\beta})_{xy}$.

So, in this case, an increase in tensor rank improves efficiency.

Comment. Examples when an increase in tensor rank is beneficial are well known in physics: e.g., a representation of a rank-1 vector as a rank-2 spinor works in relativistic quantum physics [6].

Quantum computing as computing with tensors. Classical computation is based on the idea of a *bit*: a system with two states 0 and 1. In quantum physics, due to the superposition principle, we can have states

$$c_0 \cdot |0\rangle + c_1 \cdot |1\rangle \quad (5)$$

with complex values c_0 and c_1 ; such states are called *quantum bits*, or *qubits*, for short.

The meaning of the coefficients c_0 and c_1 is that they describe the probabilities to measure 0 and 1 in the given state: $\text{Prob}(0) = |c_0|^2$ and $\text{Prob}(1) = |c_1|^2$. Because of this physical interpretations, the values c_0 and c_1 must satisfy the constraint $|c_0|^2 + |c_1|^2 = 1$.

For an n -(qu)bit system, a general state has the form

$$c_{0\dots 00} \cdot |0\dots 00\rangle + c_{0\dots 01} \cdot |0\dots 01\rangle + \dots + c_{1\dots 11} \cdot |1\dots 11\rangle. \quad (6)$$

From this description, one can see that each quantum state of an n -bit system is, in effect, a *tensor* $c_{i_1\dots i_n}$ of rank n .

In these terms, the main advantage of quantum computing is that it can enable us to store the entire tensor in only n (qu)bits. This advantage explains the known efficiency of quantum computing. For example:

- we can search in an unsorted list of n elements in time $\sqrt{n}$ — which is much faster than the time n which is needed on non-quantum computers [8, 9, 15];
- we can factor a large integer in time which does not exceed a polynomial of the length of this integer — and thus, we can break most existing cryptographic codes like widely used RSA codes which are based on the difficulty of such a factorization on non-quantum computers [15, 18, 19].

Tensors to describe constraints. A general constraint between n real-valued quantities is a subset $S \subseteq R^n$. A natural idea is to represent this subset block-by-block — by enumerating sub-blocks that contain elements of S .

Each block $b_{i_1\dots i_n}$ can be described by n indices $i_1, \dots, i_n$. Thus, we can describe a constraint by a boolean-valued tensor $t_{i_1\dots i_n}$ for which:

- $t_{i_1 \dots i_n}$ = “true” if $b_{i_1 \dots i_n} \cap S \neq \emptyset$; and
- $t_{i_1 \dots i_n}$ = “false” if $b_{i_1 \dots i_n} \cap S = \emptyset$.

Processing such constraint-related sets can also be naturally described in tensor terms.

This representation speeds up computations; see, e.g., [3,4].

Computing with tensors can also help physics. So far, we have shown that tensors can help computing. It is possible that the relation between tensors and computing can also help physics.

As an example, let us consider Kaluza-Klein-type high-dimensional space-time models of modern physics; see, e.g., [7, 11–13, 16, 20]. Einstein’s original idea [5] was to use “tensors” with integer or circular values to describe these models. From the mathematical viewpoint, such “tensors” are unusual. However, in computer terms, integer or circular data types are very natural: e.g., circular data type means fixed point numbers in which the overflow bits are ignored. Actually, from the computer viewpoint, integers and circular data are even more efficient to process than standard real numbers.

Remaining open problem. One area where tensors naturally appear is an efficient Taylor series approach to uncertainty propagation; see, e.g., [1, 14, 17]. Specifically, the dependence of the result y on the inputs $x_1, \dots, x_n$ is approximated by the Taylor series:

$$y = c_0 + \sum_{i=1}^n c_i \cdot x_i + \sum_{i=1}^n \sum_{j=1}^n c_{ij} \cdot x_i \cdot x_j + \dots \quad (7)$$

The resulting tensors $c_{i_1 \dots i_r}$ are symmetric:

$$c_{i_1 \dots i_r} = c_{\pi(i_1) \dots \pi(i_r)} \quad (8)$$

for each permutation π . As a result, the standard computer representation leads to a $r!$ duplication. An important problem is how to decrease this duplication.

3. How to Store Tensors in Computer Memory

Need to store values in computer memory. The computer memory is 1-D, so whatever multi-dimensional object we describe, its components are stored sequentially. What is the best way to arrange 2-D and higher-dimensional data in a computer memory?

Storing 2-D values in computer memory: towards formalization of the problem. Let us describe this problem in precise terms. We will start this description with the simplest case of 2-D objects.

Storing 2-D object, with components a_{ij} , $1 \leq i, j \leq n$, means assigning, to each pair (i, j) , the cell number $f(i, j)$ in such a way that different pairs (i, j) correspond to different cell numbers $f(i, j)$.

So, to describe a storing arrangement, we must describe a function

$$f : \{1, 2, \dots, n\} \times \{1, 2, \dots, n\} \rightarrow N \quad (9)$$

that maps each pair of integers $i, j \in \{1, 2, \dots, n\}$ into a natural number.

How to gauge the quality of a memory arrangement? motivations. It is desirable to arrange the storage in such a way that neighboring elements of a 2-D object are located in the memory as close to each other as possible. Neighboring elements are elements (i, j) and (i', j') for which $|i - i'| \leq 1$ and $|j - j'| \leq 1$. Thus, we can gauge the quality of the memory arrangement by the largest distance between the locations of neighboring points.

As a result, we arrive at the following numerical characteristics of the quality of different memory arrangements f .

How to gauge the quality of a memory arrangement? a formula. The quality of a memory arrangement f is described by the value

$$C(f) \stackrel{\text{def}}{=} \max\{|f(i, j) - f(i', j')| : |i - i'| \leq 1, |j - j'| \leq 1\}. \quad (10)$$

The smaller this value, the better. Thus, we are interested in finding the arrangement with the smallest possible value of the quantity $C(f)$.

Standard memory arrangement. Before we start analyzing possible memory arrangements, let us recall the standard one. In the standard programming arrangement of a 2-D array, the values are stored row by row:

- first, we have elements of the first row,

$$f(1, 1) = 1, f(1, 2) = 2, \dots, f(1, n) = n; \quad (11)$$

- then, we have elements of the second row,

$$f(2, 1) = n + 1, f(2, 2) = n + 2, \dots, f(2, n) = n + n = 2n; \quad (12)$$

- ...

- the elements of the k -th row are store at

$$\begin{aligned} f(k, 1) &= (k - 1) \cdot n + 1, f(k, 2) = (k - 1) \cdot n + 2, \dots, \\ f(k, n) &= (k - 1) \cdot n + n = k \cdot n; \end{aligned} \quad (13)$$

- ...

- finally, the elements of the last (n -th) row are stored at locations

$$\begin{aligned} f(n, 1) &= (n - 1) \cdot n + 1, f(n, 2) = (n - 1) \cdot n + 2, \dots, \\ f(n, n) &= (n - 1) \cdot n + n = n^2. \end{aligned} \quad (14)$$

Quality of the standard memory arrangement. What is the value of the quantity $C(f)$ for the standard memory arrangement f ? In other words, how far away from each other can neighboring elements (i, j) and (i', j') be located in the computer memory?

If these two elements are in the same row, i.e., if $i = i'$, then these neighboring elements (i, j) and (i, j') , with $|j - j'| = 1$, are neighbors in the memory as well:

$$|f(i, j) - f(i, j')| = |j - j'| = 1. \quad (15)$$

If these two elements are in the neighboring rows, $|i - i'| = 1$ and $|j - j'| \leq 1$, then we get

$$\begin{aligned} f(i, j) - f(i', j') &= ((i - 1) \cdot n + j) - ((i' - 1) \cdot n + j') = \\ &= (i - i') \cdot n + (j - j'). \end{aligned} \quad (16)$$

Here,

$$\begin{aligned} |f(i, j) - f(i', j')| &= |(i - i') \cdot n + (j - j')| = \\ &= |n + (j - j')| \leq n + |j - j'| \leq n + 1. \end{aligned} \quad (17)$$

Thus, for the standard memory arrangement f , the largest distance $C(f)$ between the memory locations of neighboring values cannot exceed $n + 1$. The distance between the locations of the neighboring values can be actually equal to $n + 1$: e.g., for values $(1, 1)$ and $(2, 2)$. Thus, for the standard memory arrangement, we have $C(f) = n + 1$.

A surprising result: the standard memory arrangement is optimal. Based on the fact that other memory arrangements of 2-D objects are often beneficial, one would expect these other memory arrangements be better than the standard one in the sense of our criterion $C(f)$. Surprisingly, this is not the case: it turns out that the standard memory arrangement is optimal.

To be more precise, we will prove that for every possible memory arrangement F , we have $C(f) \geq n + 1$. Thus, the standard arrangement, for which $C(f) = n + 1$, is indeed optimal.

Proof. Let us prove the inequality $C(f) \geq n + 1$. Let f be an arbitrary memory arrangement. This arrangement results in n^2 locations $f(i, j)$ corresponding to n^2 different pairs (i, j) .

Let us denote the smallest of these n^2 values by $\underline{f}$, and the largest of these values by $\overline{f}$:

$$\underline{f} \stackrel{\text{def}}{=} \min\{f(i, j) : 1 \leq i, j \leq n\}, \quad (18)$$

$$\overline{f} \stackrel{\text{def}}{=} \max\{f(i, j) : 1 \leq i, j \leq n\}. \quad (19)$$

Between $\underline{f}$ and $\overline{f}$ (including both), there are n^2 different integers. For every $a < b$, the list $a, a + 1, \dots, b$ contains $b - a + 1$ integers. Thus, we must have $\overline{f} - \underline{f} + 1 \geq n^2$, hence

$$\overline{f} - \underline{f} \geq n^2 - 1. \quad (20)$$

Let $(\underline{i}, \underline{j})$ denote the pair for which $f(\underline{i}, \underline{j}) = \underline{f}$, and let $(\overline{i}, \overline{j})$ denote the pair for which $f(\overline{i}, \overline{j}) = \overline{f}$. We can now design a sequence of pairs (i_k, j_k) going from $(i_0, j_0) = (\underline{i}, \underline{j})$ to $(i_N, j_N) = (\overline{i}, \overline{j})$ in such a way that for every k , the pairs (i_k, j_k) and (i_{k+1}, j_{k+1}) are neighbors.

Indeed, if $\underline{i} < \bar{i}$, we start with $i_0 = \underline{i}$, and then take $i_1 = i_0 + 1$, $i_2 = i_0 + 2$, etc., until we reach $\bar{i}$ — after this, we continue to take $i_k = \bar{i}$.

If $\underline{i} > \bar{i}$, we start with $i_0 = \underline{i}$, and then take $i_1 = i_0 - 1$, $i_2 = i_0 - 2$, etc., until we reach $\bar{i}$ — after this, we continue to take $i_k = \bar{i}$.

If $\underline{i} = \bar{i}$, then we simply take $i_k = \underline{i}$ for all k .

Similarly, if $\underline{j} < \bar{j}$, we start with $j_0 = \underline{j}$, and then take $j_1 = j_0 + 1$, $j_2 = j_0 + 2$, etc., until we reach $\bar{j}$ — after this, we continue to take $j_k = \bar{j}$.

If $\underline{j} > \bar{j}$, we start with $j_0 = \underline{j}$, and then take $j_1 = j_0 - 1$, $j_2 = j_0 - 2$, etc., until we reach $\bar{j}$ — after this, we continue to take $j_k = \bar{j}$.

If $\underline{j} = \bar{j}$, then we simply take $j_k = \underline{j}$ for all k .

At each step, each of the coordinates is changed by at most 1, so the pairs (i_k, j_k) and (i_{k+1}, j_{k+1}) are indeed neighbors.

We need $|\underline{i} - \bar{i}| + 1$ steps to reach from $\underline{i}$ to $\bar{i}$, and we need $|\underline{j} - \bar{j}|$ steps to reach from $\underline{j}$ to $\bar{j}$. Thus, overall, we need

$$N = \max(|\underline{i} - \bar{i}|, |\underline{j} - \bar{j}|) + 1 \quad (21)$$

steps. For values from 1 to n , the largest possible difference $|\underline{j} - \bar{j}|$ is equal to $n - 1$, hence $N \leq n$.

Now, we have

$$\begin{aligned} f(\underline{i}, \underline{j}) - f(\bar{i}, \bar{j}) &= f(i_0, j_0) - f(i_N, j_N) = \\ &= (f(i_0, j_0) - f(i_1, j_1)) + (f(i_1, j_1) - f(i_2, j_2)) + \dots + \\ &\quad + (f(i_{N-1}, j_{N-1}) - f(i_N, j_N)). \end{aligned} \quad (22)$$

Thus,

$$\begin{aligned} |f(\underline{i}, \underline{j}) - f(\bar{i}, \bar{j})| &\leq \\ &\leq |f(i_0, j_0) - f(i_1, j_1)| + |f(i_1, j_1) - f(i_2, j_2)| + \dots + \\ &\quad + |f(i_{N-1}, j_{N-1}) - f(i_N, j_N)|. \end{aligned} \quad (23)$$

Since for each k , the pairs (i_k, j_k) and (i_{k+1}, j_{k+1}) are neighbors, we have

$$|f((i_k, j_k) - f(i_{k+1}, j_{k+1}))| \leq C(f).$$

So, from (23), we conclude that

$$|f(\underline{i}, \underline{j}) - f(\bar{i}, \bar{j})| \leq (N - 1) \cdot C(f). \quad (24)$$

Since $N \leq n$, we thus have

$$|\underline{f} - \bar{f}| = |f(\underline{i}, \underline{j}) - f(\bar{i}, \bar{j})| \leq (n - 1) \cdot C(f). \quad (25)$$

On the other hand, we know that $n^2 - 1 \leq |\underline{f} - \bar{f}|$. Thus, we conclude that

$$n^2 - 1 \leq (n - 1) \cdot C(f), \quad (26)$$

and therefore, that

$$C(f) \geq \frac{n^2 - 1}{n - 1} = n + 1. \quad (27)$$

The statement is proven.

The standard memory arrangement is not the only optimal one. The fact that the standard memory arrangement turned out to have the optimal (smallest possible) value of $C(f)$ may not sound so surprising if we realize that several different memory arrangements have the exact same optimal value of $C(f)$.

One such arrangement is clear: instead of storing the values row by row, we can store them column by column:

- first, we have elements of the first column,

$$f(1, 1) = 1, f(2, 1) = 2, \dots, f(n, 1) = n; \quad (28)$$

- then, we have elements of the second column,

$$f(1, 2) = n + 1, f(2, 2) = n + 2, \dots, f(n, 2) = 2n; \quad (29)$$

- ...

- the elements of the k -th column are store at

$$\begin{aligned} f(1, k) &= (k - 1) \cdot n + 1, f(2, k) = (k - 1) \cdot n + 2, \dots, \\ f(n, k) &= (k - 1) \cdot n + n = k \cdot n; \end{aligned} \quad (30)$$

- ...

- finally, the elements of the last (n -th) column are stored at locations

$$\begin{aligned} f(1, n) &= (n - 1) \cdot n + 1, f(2, n) = (n - 1) \cdot n + 2, \dots, \\ f(n, n) &= (n - 1) \cdot n + n = n^2. \end{aligned} \quad (31)$$

There are other examples as well: e.g., elements of a 2×2 matrix can be stored in the order $(1, 1), (1, 2), (2, 2), (2, 1)$ with the same value $C(f) = n + 1 = 3$ as for row-by-row or column-by-column memory arrangements.

Multi-dimensional case. In the k -dimensional case, we need to assign location $f(i_1, \dots, i_k)$ to tuples $(i_1, \dots, i_k)$. It is also natural to gauge the quality of the memory arrangement by the largest distance between the locations of neighboring values, i.e., tuples $(i_1, \dots, i_k)$ and $(i'_1, \dots, i'_k)$ for which $|i_j - i'_j| \leq 1$ for all j . The quality of a memory arrangement f can be thus naturally described by the value

$$\begin{aligned} C(f) &\stackrel{\text{def}}{=} \max\{|f(i_1, \dots, i_k) - f(i'_1, \dots, i'_k)| : \\ &|i_j - i'_j| \leq 1 \text{ for all } j = 1, \dots, k\}. \end{aligned} \quad (32)$$

In the standard computer arrangement, we store elements in lexicographic order: i.e., $(i_1, \dots, i_k)$ is placed before $(i'_1, \dots, i'_k)$ if for the first differing coordinate $i_j \neq i'_j$, we have $i_j < i'_j$. In other words, we first store values

$$(1, \dots, 1), \dots, (1, \dots, n), \quad (33)$$

then values

$$(1, \dots, 2, 1), \dots, (1, \dots, 2, n), \quad (34)$$

etc. In this arrangement,

- the difference in the last coordinate $i_k - i'_k = 1$ leads to a difference of 1 in memory locations;
- the difference in the next to last coordinate $i_{k-1} - i'_{k-1} = 1$ leads to a difference of n in memory locations,
- $\dots$,
- the difference in the first coordinate $i_1 - i'_1$ leads to a difference of n^{k-1} in memory locations.

Thus, the difference in location of neighboring tuples cannot exceed

$$n^{k-1} + n^{k-2} + \dots + n + 1.$$

This distance is attained, e.g., for the points $(1, \dots, 1)$ and $(2, \dots, 2)$. Thus, for the standard memory arrangement f , we have

$$C(f) = n^{k-1} + n^{k-2} + \dots + n + 1. \quad (35)$$

Similarly to the 2-D case, we can prove that this memory arrangement is optimal. Indeed, in this case, for the difference between the values

$$\underline{f} \stackrel{\text{def}}{=} \min\{f(i_1, \dots, i_k) : 1 \leq i_j \leq n\}, \quad (36)$$

$$\overline{f} \stackrel{\text{def}}{=} \max\{f(i_1, \dots, i_k) : 1 \leq i_j \leq n\}, \quad (37)$$

we have $\overline{f} - \underline{f} \geq n^k - 1$. We can still move from the tuple $(\underline{i}_1, \dots, \underline{i}_k)$ at which the smallest value $\underline{f}$ is attained to the tuple $(\overline{i}_1, \dots, \overline{i}_k)$ at which the largest value $\overline{f}$ is attained in $\leq n - 1$ transitions from a tuple to a neighboring one. Thus, we can conclude that

$$n^k - 1 \leq (n - 1) \cdot C(f), \quad (38)$$

and therefore, that

$$C(f) \geq \frac{n^k - 1}{n - 1} = n^{k-1} + n^{k-2} + \dots + n + 1. \quad (39)$$

The optimality is proven.

Acknowledgment

This work was supported in part by the US National Science Foundation grant HRD-1242122 (Cyber-ShARE Center of Excellence).

The authors are thankful to Fred G. Gustavson and Lenore Mullin for her encouragement.

REFERENCES

1. Berz M., Hoffstätter G. Computation and application of Taylor polynomials with interval remainder bounds // *Reliable Computing*. 1998. No. 4(1). P. 83–97.
2. Bryant R.E., O'Hallaron D.R. *Computer Systems: A Programmer's Perspective*. Upper Saddle River : Prentice Hall, 2003.
3. Ceberio M., Ferson S., Kreinovich V., Chopra S., Xiang G., Murguía A., Santillan J. How to take into account dependence between the inputs: from interval computations to constraint-related set computations, with potential applications to nuclear safety, bio- and geosciences // *Journal of Uncertain Systems*. 2007. No. 1(1). P. 11–34.
4. Ceberio M., Kreinovich V., Pownuk A., Bede B. From interval computations to constraint-related set computations: towards faster estimation of statistics and ODEs under interval, p-box, and fuzzy uncertainty // *Foundations of Fuzzy Logic and Soft Computing* / P. Melin, O. Castillo, L.T. Aguilar, J. Kacprzyk, W. Pedrycz (eds.). Proceedings of the World Congress of the International Fuzzy Systems Association IFSA'2007. Cancun, Mexico, June 18–21, 2007. Springer Lecture Notes on Artificial Intelligence. 2007. No. 4529. P. 33–42.
5. Einstein A., Bergmann P. On the generalization of Kaluza's theory of electricity // *Ann. Phys.* 1938. No. 39. P. 683–701.
6. Feynman R., Leighton R., Sands M. *The Feynman Lectures on Physics*. Boston : Addison Wesley, 2005.
7. Green M.B., Schwarz J.H., Witten E. *Superstring Theory*. Vol. 1–2. Cambridge University Press, 1988.
8. Grover L.K. A fast quantum mechanical algorithm for database search // *Proceedings of the 28th Annual ACM Symposium on the Theory of Computing*. May 1996. P. 212–ff.
9. Grover L.K. From Schrödinger's equation to quantum search algorithm // *American Journal of Physics*. 2001. No. 69(7). P. 769–777.
10. Gustavson F.G. The relevance of new data structure approaches for dense linear algebra in the new multi-core/many core environments // *Proceedings of the 7th International Conference on Parallel Processing and Applied Mathematics PPAM'2007*. Gdansk, Poland, September 9–12, 2007. Springer Lecture Notes in Computer Science. 2008. No. 4967. P. 618–621.
11. Kaluza Th. *Sitzungsberichte der K. Preussischen Akademie der Wissenschaften zu Berlin*. 1921. P. 966 (in German); Engl. translation: On the unification problem in physics [13, p. 1–9].
12. Klein O. *Zeitschrift für Physik*. 1926. Vol. 37. P. 895 (in German); Engl. translation: Quantum theory and five-dimensional relativity [13, p. 10–23].
13. Lee H.C. (ed.). *An Introduction to Kaluza-Klein Theories*. Singapore : World Scientific, 1984.
14. Neumaier A. Taylor forms // *Reliable Computing*. 2002. No. 9. P. 43–79.
15. Nielsen M., Chuang I. *Quantum Computation and Quantum Information*. Cambridge : Cambridge University Press, 2000.
16. Polchinski J. *String Theory*. V. 1–2. Cambridge University Press, 1998.
17. Revol N., Makino K., Berz M. Taylor models and floating-point arithmetic: proof that arithmetic operations are validated in COSY // *J. Log. Algebr. Program.* 2005. No. 64(1). P. 135–154.

18. Shor P. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer // Proceedings of the 35th Annual Symposium on Foundations of Computer Science. Santa Fe, NM, Nov. 20–22, 1994.
19. Shor P. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer // SIAM J. Sci. Statist. Comput. 1997. V. 26. P. 1484–ff.
20. Starks S.A., Kosheleva O., Kreinovich V. Kaluza-Klein 5D ideas made fully geometric // International Journal of Theoretical Physics. 2006. No. 45(3). P. 589–601.
21. Tietze H. Famous Problems of Mathematics: Solved and Unsolved Mathematical Problems, from Antiquity to Modern Times. New York : Graylock Press, 1965.
22. Zaniolo C., Ceri S., Faloutsos C., Snodgrass R.T., Subrahmanian V.S., Zicari R. Advanced Database Systems. Morgan Kaufmann, 1997.

КАК ХРАНИТЬ ТЕНЗОРЫ В ПАМЯТИ КОМПЬЮТЕРА: ОБЗОР

М. Себерьо

к.ф.-м.н., доцент, e-mail: mceberio@utep.edu

В. Крейнович

к.ф.-м.н., профессор, e-mail: vladik@utep.edu

Техасский университет в Эль Пасо, США

Аннотация. В этой статье, объяснив необходимость использования тензоров в вычислениях, мы анализируем вопрос о том, как лучше хранить тензоры в памяти компьютера. Оказывается, что относительно естественного критерия оптимальности стандартный способ хранения тензоров оказывается одним из оптимальных.

Ключевые слова: Тензоры, вычислительная техника, память компьютера.

Дата поступления в редакцию: 02.01.2018

HOW INTERVAL MEASUREMENT UNCERTAINTY AFFECTS THE RESULTS OF DATA PROCESSING: A CALCULUS-BASED APPROACH TO COMPUTING THE RANGE OF A BOX

Andrew Pownuk

Ph.D. (Phys.-Math.), e-mail: ampownuk@utep.edu

Vladik Kreinovich

Ph.D. (Phys.-Math.), Professor, e-mail: vladik@utep.edu

University of Texas at El Paso, El Paso, Texas 79968, USA

Abstract. In many practical applications, we are interested in the values of the quantities $y_1, \dots, y_m$ which are difficult (or even impossible) to measure directly. A natural idea to estimate these values is to find easier-to-measure related quantities $x_1, \dots, x_n$ and to use the known relation to estimate the desired values y_j . Measurements come with uncertainty, and often, the only thing we know about the actual value of each auxiliary quantity x_i is that it belongs to the interval $[\underline{x}_i, \bar{x}_i] = [\tilde{x}_i - \Delta_i, \tilde{x}_i + \Delta_i]$, where $\tilde{x}_i$ is the measurement result, and Δ_i is the upper bound on the absolute value of the measurement error $\tilde{x}_i - x_i$. In such situations, instead of a single value of a tuple $y = (y_1, \dots, y_m)$, we have a range of possible values. In this paper, we provide calculus-based algorithms for computing this range.

Keywords: Data processing, interval uncertainty, indirect measurements, calculus.

1. Formulation of the Problem

Need for indirect measurements. In many practical situations, we are interested in the values of the quantities $y_1, \dots, y_m$ which are difficult — or even impossible — to measure directly. Since we cannot measure these quantities directly, a natural idea is to measure them *indirectly* (see, e.g., [6]), i.e.:

- to measure related quantities $x_1, \dots, x_n$ which are related to the desired quantities y_j by known relations, and
- to use appropriate algorithms to find the values of the desired quantities:

$$\begin{aligned} y_1 &= f_1(x_1, \dots, x_n); \\ y_2 &= f_2(x_1, \dots, x_n); \\ &\dots \\ y_m &= f_m(x_1, \dots, x_n). \end{aligned} \tag{1}$$

Comment. In the real world, the relations are usually smooth; see, e.g., [1, 7].

Need to take into account measurement uncertainty. If we knew the exact values $x_1, \dots, x_n$ of all the auxiliary quantities, then, by using the relations (1), we would be able to find the exact values of all the desired quantities $y_1, \dots, y_m$.

In practice, however, measurements are never absolutely precise. The measurement result $\tilde{x}_i$ is, in general, different from the actual (unknown) values of the corresponding quantity. When we plus in values $\tilde{x}_i \neq x_i$ into the formula (1), we, in general, get the values $\tilde{y}_j = f_j(\tilde{x}_1, \dots, \tilde{x}_n)$ which are, in general, different from the desired values y_j . How can we gauge the resulting uncertainty in y_j ?

Case of interval measurement uncertainty. In many practical situations, the only information that we have about the measurement error $\Delta x_i \stackrel{\text{def}}{=} \tilde{x}_i - x_i$ is the upper bound Δ_i provided by the manufacturer of the corresponding measuring instrument. (If the manufacturer provide no such bound, then it is not a measuring instrument, it is a device for producing wild guesses.)

In this case, once we know the measurement result $\tilde{x}_i$, the only information we have about the actual value x_i is that it is somewhere on the interval $[\underline{x}_i, \bar{x}_i]$, where $\underline{x}_i \stackrel{\text{def}}{=} \tilde{x}_i - \Delta_i$ and $\bar{x}_i \stackrel{\text{def}}{=} \tilde{x}_i + \Delta_i$; see, e.g., [2, 4–6].

There is no a priori known relation between the values x_i , so the set of all possible values of x_i should not depend on the values of all other quantities x_j , $j \neq i$. Thus, the set of all possible values of the tuple $x = (x_1, \dots, x_n)$ is the box

$$[\underline{x}_1, \bar{x}_1] \times \dots \times [\underline{x}_n, \bar{x}_n]. \quad (2)$$

Resulting problem. Once we know that x belongs to the box (2), what are the possible values of the tuple $y = (y_1, \dots, y_m)$? In mathematical terms, what is the range of the box (2) under the mapping (1)?

In this paper, we describe calculus-based techniques for solving this problem.

2. Analysis of the Problem and the Resulting Algorithms

Simplest case when we have only one desired quantity y_1 : analysis of the problem. Let us start with the simplest case, when we have only one desired quantity y_1 . In this case, we are interested in the range of the function $f_1(x_1, \dots, x_n)$ when each x_i is in the corresponding interval $[\underline{x}_i, \bar{x}_i]$. For smooth (even for continuous) functions, this range is connected and is, thus, an interval $[\underline{y}_1, \bar{y}_1]$, where:

- $\underline{y}_1$ is the smallest possible value of the function $f_1(x_1, \dots, x_n)$ on the given box, and
- $\bar{y}_1$ is the largest possible value of the function $f_1(x_1, \dots, x_n)$ on the given box.

For each variable x_i , the maximum (or minimum) of the expression $y_1 = f_1(x, \dots, x_n)$ is attained:

- either at one of the endpoints of this interval, i.e., for $x_i = \underline{x}_i$ or $x_i = \bar{x}_i$,
- or inside the corresponding interval $(\underline{x}_i, \bar{x}_i)$.

According to calculus, if the maximum or minimum is attained inside an interval, then the corresponding derivative $\frac{\partial f_1}{\partial x_i}$ is equal to 0. So, for each i , it is sufficient to consider three possible cases:

- the case when $x_i = \underline{x}_i$;
- the case when $x_i = \bar{x}_i$, and
- the case when $\frac{\partial f_1}{\partial x_i} = 0$.

Thus, to find the minimum $\underline{y}_1$ and the maximum $\bar{y}_1$ of the function $y_1 = f_1(x_1, \dots, x_n)$ over the box, it is sufficient to consider all possible combinations of these 3 cases.

In other words, we arrive at the following algorithm.

Case when we have only one desired quantity y_1 : algorithm. Consider all systems of equations, in which, for each i , we have one of the three alternatives: $x_i = \underline{x}_i$, $x_i = \bar{x}_i$, and $\frac{\partial f_1}{\partial x_i} = 0$. There are 3^n such systems.

For each of these systems, we find the corresponding values $x = (x_1, \dots, x_n)$ and compute the corresponding value $y_1 = f(x_1, \dots, x_n)$. The largest of thus computed values is $\bar{y}_1$, the smallest is $\underline{y}_1$.

Comment. This algorithm requires solving an exponential number of systems and thus takes exponential time. This is, however, unavoidable, since it is known that already for quadratic functions $f_1(x_1, \dots, x_n)$, the problem of computing the bounds $\underline{y}$ and $\bar{y}$ is NP-hard; see, e.g., [3]. This means that, unless $P=NP$ (which most computer scientists believe to be impossible), super-polynomial (e.g., exponential) computation time is unavoidable — at least for some inputs.

Exponential time does not mean that the algorithm is not practical — for reasonably small n , solving 3^n system is quite reasonable. For example, for $n = 10$, we need to solve less than 60,000 systems, it is a large number, but it is quite doable. For $n = 15$, we need to solve about 5 million systems — still possible.

What we plan to do next. In the following subsections, we show how we can extend this calculus-based approach to the general case, and thus reduce the difficult-to-solve problem of finding the range to more well-studied problems of solving systems of equations.

Case when the number m of desired quantities is equal to the number n of auxiliary ones: analysis of the problem. To find the range means to find its border. At almost all points on the border, there is — locally — at least one tangent plane. A plane in an m -dimensional space has the form

$$\sum_{j=1}^m c_j \cdot y_j = c_0.$$

Thus, at this border point $y = (y_1, \dots, y_m)$, the linear expression

$$y = \sum_{j=1}^m c_j \cdot y_j = f(x_1, \dots, x_n) \stackrel{\text{def}}{=} \sum_{j=1}^m c_j \cdot f_j(x_1, \dots, x_n)$$

attains its local maximum or local minimum.

Similarly to the previous case, this may mean that one of the inputs x_i attains its largest possible value $\bar{x}_i$ or its smallest possible value $x_i = \underline{x}_i$. In this case, the corresponding condition $x_i = \underline{x}_i$ or $x_i = \bar{x}_i$ determines the $(n - 1)$ -dimensional set – which could be part of the border.

It may also mean that the maximum or minimum of the linear function is attained when all the values x_i are inside the corresponding intervals. In this case, we get

$$\frac{\partial f}{\partial x_i} = 0$$

for all i , i.e., we get

$$\sum_{j=1}^m c_j \cdot \frac{\partial f_j}{\partial x_i} = 0$$

for all i .

In algebraic terms, the existence of non-always-zero values c_j that satisfy the above equality for all i means that $m = n$ gradient vectors

$$\left(\frac{\partial f_j}{\partial x_1}, \dots, \frac{\partial f_j}{\partial x_n} \right)$$

that correspond to different j are linearly dependent. According to linear algebra, this is equivalent to requiring that the determinant of the Jacobian matrix $\left\| \frac{\partial f_j}{\partial x_i} \right\|$ is equal to 0:

$$\det \left\| \frac{\partial f_j}{\partial x_i} \right\| = 0. \quad (3)$$

So, we arrive at the following algorithm.

Case when the number m of desired quantities is equal to the number n of auxiliary ones: algorithm. To find the border of the desired range, for each i from 1 to $m = n$, we form two systems of equations:

- the system (1) in which we substitute $x_i = \underline{x}_i$, and
- the system (1) in which we substitute $x_i = \bar{x}_i$.

Each of these systems provides a set of co-dimension 1 that could potentially serve as part of the border of the desired set.

To these possible border sets, we add the set corresponding to the equation (3). This equation defined a set of co-dimension 1, and plugging this set into (1), we can a y -set of co-dimension one – which can also be part of the border.

We know that the actual border can contain only segments of the above type, so once we have computed all these segments, we can reconstruct the border.

General case: analysis of the problem. We have already considered the case when $m = n$. There are two remaining cases: when $n < m$ and when $m < n$.

When $n < m$, the set of all possible values of the tuple y is of smaller dimension than the m , so this set is its own boundary.

Let us now consider the case when $m < n$. In this case, also, some linear combination

$$f(x_1, \dots, x_n) = \sum_{j=1}^m c_j \cdot f_j(x_1, \dots, x_n)$$

attains its maximum or its minimum. Let v denote the number of inputs x_i for which at this maximum-or-minimum point, we have $x_i = \underline{x}_i$ or $x_i = \bar{x}_i$. For each of the remaining $n - v$ variables x_i , we then have the equation

$$\sum_{j=1}^m c_j \cdot \frac{\partial f_j}{\partial x_i} = 0. \quad (4)$$

This equality (4) must hold for all $(n - v)$ values of i , so we must have $(n - v)$ equations.

We can select one of the values c_j equal to 1, then the other $m - 1$ values of c_j can be determining if we consider the first $m - 1$ conditions (4) as a system of linear equations with $m - 1$ unknowns. Substituting these values for c_j into the remaining $n - v - (m - 1)$ equalities (4), we thus get $n - v - (m - 1)$ equalities that relate $n - v$ unknowns.

In general, each additional equality imposed on elements of a set decreases its dimension by 1. For example, in the 3-D space:

- the set of all the points that satisfy a certain equality is usually a 2-D surface,
- the set of points that satisfy two independent equalities is a 1-D line, etc.

In our case, the dimension of the set of all the $(n - v)$ -dimensional tuples x that satisfy all $n - v - (m - 1)$ equalities is equal to the difference

$$(n - v) - (n - v - (m - 1)) = m - 1.$$

The image of this $(m - 1)$ -dimensional set under the transformation (1) is also $(m - 1)$ -dimensional, so it forms a surface in the m -dimensional space of all possible tuples $y = (y_1, \dots, y_m)$.

As a result, we get the following algorithm.

General case: algorithm. We consider all possible subsets I of the set $\{1, \dots, n\}$ of all indices of the inputs x_i . For each such subset I of size v , we consider all 2^v possible combinations of values $\underline{x}_i$ and $\bar{x}_i$.

For each such combination, we consider the system of equations (4) for all $i \notin I$. We can set up one of the values c_j to 1 and the first $m - 1$ equations (4) to describe c_j as a function of $x_1, \dots, x_m$. Substituting the resulting expressions for c_j in terms of x_i into the remaining $n - v - (m - 1)$ equalities (4), we get a $(m - 1)$ -dimensional set of tuples x . Substituting this set of tuples into the formula (1), we get a $(m - 1)$ -dimensional set of y -tuples.

We thus get several $(m - 1)$ -dimensional sets, and we know that the actual border can only consist of the above fragments.

Acknowledgments

This work was supported in part by the National Science Foundation grant HRD-1242122 (Cyber-ShARE Center of Excellence).

The authors are thankful to all the participants of the 2017 UTEP/NMSU Workshop on Mathematics, Computer Science, and Computation Science (El Paso, Texas, November 4, 2017) for valuable discussions.

REFERENCES

1. Feynman R., Leighton R., Sands M. The Feynman Lectures on Physics. Addison Wesley, Boston, Massachusetts, 2005.
2. Jaulin L., Kiefer M., Didrit O., Walter E. Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control, and Robotics. Springer, London, 2001.
3. Kreinovich V., Lakeyev A., Rohn J., Kahl P. Computational Complexity and Feasibility of Data Processing and Interval Computations. Kluwer, Dordrecht, 1998.
4. Mayer G. Interval Analysis and Automatic Result Verification. De Gruyter, Berlin, 2017.
5. Moore R.E., Kearfott R.B., Cloud M.J. Introduction to Interval Analysis. SIAM, Philadelphia, 2009.
6. Rabinovich S.G. Measurement Errors and Uncertainty: Theory and Practice. Springer Verlag, Berlin, 2005.
7. Thorne K.S., Blandford R.D. Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics. Princeton University Press, Princeton, New Jersey, 2017.

КАК НЕОПРЕДЕЛЁННОСТЬ ИЗМЕРЕНИЯ ИНТЕРВАЛА ВЛИЯЕТ НА РЕЗУЛЬТАТЫ ОБРАБОТКИ ДАННЫХ: ВЫЧИСЛИТЕЛЬНЫЙ ПОДХОД ДЛЯ РАСЧЁТА ОБЛАСТИ ЗНАЧЕНИЙ ПРЯМОУГОЛЬНИКА

А. Повнук

к.ф.-м.н., ст. преподаватель, e-mail: amponuk@utep.edu

В. Крейнович

к.ф.-м.н., профессор, e-mail: vladik@utep.edu

Техасский университет в Эль Пасо, США

Аннотация. Во многих практических приложениях нас интересуют значения величин $y_1, \dots, y_m$, которые трудно (или даже невозможно) измерить непосредственно. Естественная идея оценить эти значения — найти более лёгкие для оценки величины $x_1, \dots, x_n$ и использовать известное отношение для оценки желаемых значений y_j . Измерения проходят с неопределённостью, и часто единственное, что мы знаем о фактическом значении каждой вспомогательной величины x_i , — это то, что оно принадлежит интервалу $[x_i, \bar{x}_i] = [\tilde{x}_i - \Delta_i, \tilde{x}_i + \Delta_i]$, где $\tilde{x}_i$ — результат измерения, а Δ_i — верхняя граница по абсолютной величине ошибки измерения

$\tilde{x}_i - x_i$. В таких ситуациях вместо одного значения кортежа $y = (y_1, \dots, y_m)$ мы имеем диапазон возможных значений. В этой статье мы предлагаем вычислительные алгоритмы для расчёта этого диапазона.

Ключевые слова: обработка данных, интервальная неопределённость, косвенные измерения, вычисления.

Дата поступления в редакцию: 25.04.2018

АВТОМАТИЗАЦИЯ ТЕСТИРОВАНИЯ СЕТЕВЫХ СРЕДСТВ ЗАЩИТЫ ИНФОРМАЦИИ НА ОСНОВЕ ПРИМЕНЕНИЯ ЭВОЛЮЦИОННО–ГЕНЕТИЧЕСКОГО ПОДХОДА

Н.И. Синадский

доцент, к.т.н., e-mail: nickis@el.ru

А.В. Агафонов

ассистент, e-mail: avagaf@gmail.com

Институт радиоэлектроники и информационных технологий УрФУ
им. первого Президента России Б.Н. Ельцина, Екатеринбург, Россия

Аннотация. В статье рассмотрен подход к тестированию устойчивости сетевых средств защиты информации к компьютерным атакам типа «отказ в обслуживании» на основе применения аппарата генетических алгоритмов.

Ключевые слова: тестирование, сетевые средства защиты информации, отказ в обслуживании, генетический алгоритм.

Введение

Одной из актуальных угроз информационной безопасности существующих информационно-телекоммуникационных сетей являются компьютерные атаки типа «отказ в обслуживании», направленные как на отдельные узлы данных сетей, так и на их технологическую инфраструктуру.

Объектами атак могут являться в том числе и сетевые средства защиты информации (ССЗИ), такие как межсетевые экраны и системы предотвращения вторжений. Успешная реализация атаки в данных случаях приводит к нарушению или существенному снижению доступности информации, обрабатываемой узлами сетей, что актуализирует задачу тестирования устойчивости ССЗИ от атак указанного типа.

Наиболее широко применяемым методом оценки защищённости оборудования компьютерных сетей от атак типа «отказ в обслуживании» является его натурное тестирование в изолированной сетевой среде с применением синтезированного сетевого трафика (СТ), имитирующего комбинацию СТ штатного информационного взаимодействия узлов компьютерной сети и атакующего воздействия.

Разработанная авторами методика тестирования ССЗИ подразумевает передачу тестового СТ, в процессе которой производится оценка способности тестируемого оборудования обеспечивать заданный требованиями компьютерной

сети уровень доступности информации. Данный уровень может быть описан совокупностью следующих параметров: среднего значения задержки передачи пакетов $\overline{t_d}$, его среднеквадратического отклонения $\sigma(\delta t_p)$, называемого также джиттером, и относительной доли потерь пакетов q , — которые могут быть заданы вектором в пространстве $\Omega = \{\omega\}$, где вектор ω определяется следующим выражением:

$$\omega = \langle \overline{t_d}, \sigma(\delta t_p), q \rangle. \quad (1)$$

Особенностью рассматриваемых сетевых компьютерных атак является то, что при их реализации не задействуется прикладной уровень модели OSI и, в большинстве случаев, применяется СТ, соответствующий спецификациям используемых протоколов передачи данных, параметры которого отличаются от штатного СТ лишь количественно [1].

Поэтому до проведения тестирования ССЗИ нельзя предугадать все возможные сочетания параметров СТ атакующего воздействия, к которому оно оказывается уязвимо.

Современные исследования показали влияние на успешность реализации сетевых компьютерных атак типа «отказ в обслуживании», направленных на ССЗИ, следующих параметров СТ: n_e , n_w , n_z — количества взаимодействующих узлов, сетей и задействованных в процессе взаимодействия сетевых интерфейсов ССЗИ; p_{tcp} , p_{udp} , p_{icmp} — относительных долей потоков TCP, UDP и сеансов взаимодействия ICMP; $\overline{n_f}$, $\overline{t_f}$, $\overline{l_p}$, $\overline{t_p}$, $\overline{p_{h1}}$ — средних значений количества, длительности потоков, размера пакетов, межпакетного временного интервала и относительной доли пакетов, сгенерированных узлами-инициаторами логических соединений; $\sigma(n_f)$, $\sigma(t_f)$, $\sigma(l_p)$, $\sigma(t_p)$, $\sigma(p_{h1})$ — соответствующих им среднеквадратических отклонений.

Таким образом, может быть определено пространство $\Psi = \{\psi\}$ параметров тестового СТ, значимых в задаче оценки защищённости ССЗИ от сетевых компьютерных атак типа «отказ в обслуживании», где вектор ψ определяется следующим выражением:

$$\psi = \langle n_e, n_w, n_z, p_{tcp}, p_{udp}, p_{icmp}, \overline{n_f}, \overline{t_f}, \overline{l_p}, \overline{t_p}, \overline{p_{h1}}, \sigma(n_f), \sigma(t_f), \sigma(l_p), \sigma(t_p), \sigma(p_{h1}) \rangle \quad (2)$$

Поиск сочетаний параметров СТ атакующего воздействия, к которому ССЗИ оказывается уязвимо, необходимый для обеспечения полноты тестирования, является задачей переборного типа, которая может быть сведена к задаче отыскания экстремума многомерной функции $\psi(\omega)$. Численное решение данного класса задач может быть затруднено в связи с размерностью и видом исследуемой функции, которая в общем случае может быть нелинейной, разрывной, недифференцируемой и многоэкстремальной [2].

Наиболее перспективным методом решения данного класса задач является эволюционно-генетический подход, который используется для построения алгоритмов поиска оптимальных решений, называемых генетическими алгоритмами, на основе моделирования таких механизмов биологической эволюции, как размножение, мутация и отбор особей популяции.

Генетический алгоритм тестирования сетевых средств защиты информации

Блок-схема разработанного генетического алгоритма приведена на рис. 1.

Каждая из особей μ_i популяции $\mu = \{\mu_i\}_{i=1}^{n_\mu}$ представляет собой совокупность значений статистических параметров, на основе которых производится синтез тестового СТ, имеющего заданную структуру. Параметры особи μ_i представляются в процессе работы генетического алгоритма в виде последовательности бит $\chi_i = \langle \chi_{i,j} \rangle_{j=1}^{n_\chi}$, где $\chi_i = \{0, 1\}$, называемой далее хромосомой.

На предварительном этапе работы генетического алгоритма с использованием функции $RandomM(M)$, где M — математическая модель, описывающая структуру параметров особи, производится инициализация параметров особей случайными значениями, на основе которых затем выполняется синтез образцов тестового СТ с заданными характеристиками $\psi_i \in \Psi$.

Образцы синтезированного СТ используются для тестирования ССЗИ, в процессе которого определяется уровень обеспечиваемой ССЗИ доступности информации, $\omega_i \in \Omega$. Данные векторы $\omega = \{\omega_i\}_{i=1}^{n_\omega}$ затем используются для ранжирования популяции по убыванию значений критерия оптимальности γ_i , определяемого для особи μ_i как количество особей популяции, которым соответствуют меньшие значения всех параметров доступности информации, входящих в вектор ω_i :

$$\gamma_i = \left| \bigcap_{k=1}^3 \{ \mu_k \mid \forall j : \omega_{i,j} > \omega_{k,j} \} \right|. \quad (3)$$

В разработанном генетическом алгоритме размер популяции изменяется на каждом шаге его работы на основе анализа динамики изменения максимальных и средних значений параметров доступности информации по популяции.

В случае если ни один из элементов вектора максимальных значений не увеличил в течение шага работы генетического алгоритма своего значения, то принимается гипотеза о том, что комбинации существующих решений в процессе кроссинговера не показывают большей степени приспособленности, чем существующие, поэтому для увеличения скорости поиска новых решений, не являющихся комбинацией существующих, производится увеличение численности популяции. В случае если ни один из элементов вектора средних значений не увеличил своего значения, то принимается гипотеза о том, что популяцией обнаружен и исследуется новый локальный экстремум функции $\psi(\omega)$, поэтому для увеличения быстродействия генетического алгоритма и давления отбора размер популяции уменьшается.

Шаг изменения размера популяции генетического алгоритма принимается равным трём, так как в соответствии с механизмом скрещивания перенос одной родительской особи в следующее поколение вызывает появление двух дополнительных потомков. Минимальный размер популяции равен шести, так как для процедуры скрещивания необходимы как минимум две особи, каждая из которых генерирует по два потомка.

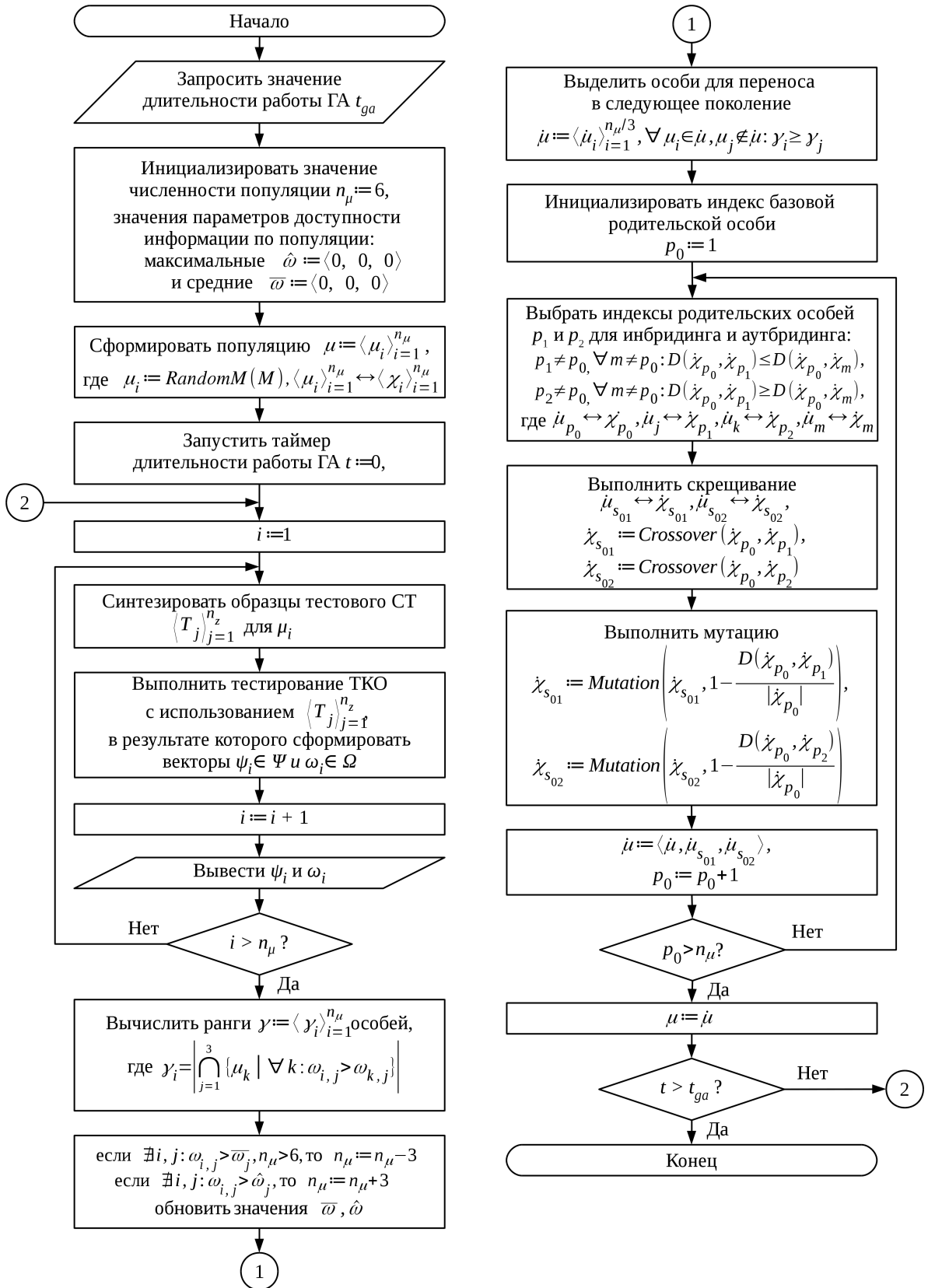


Рис. 1. Блок-схема генетического алгоритма

Для выделения особей, переходящих в следующее поколение в процессе выполнения разработанного генетического алгоритма, используется механизм элитного отбора, заключающийся в построении популяции следующего поколения из имеющих наибольшее значение критерия оптимальности.

В качестве механизма отбора особей для скрещивания используется метод, являющийся комбинацией инбридинга и аутбридинга. Инбридинг заключается в выборе пар особей популяции, имеющих наименьшие различия особей; аутбридинг — в выборе особей, имеющих наибольшие различия, мерой которых является расстояние Хэмминга $D(\chi)$ между их хромосомами.

Подбор особей в родительские пары при инбридинге приводит к скрещиванию особей со сходными параметрами, поэтому данный механизм позволяет сохранить имеющиеся удачные сочетания параметров СТ, производя поиск больших значений функции $\omega(\psi)$ вблизи родительских особей.

Аутбридинг позволяет избежать потери разнообразия исследуемых сочетаний параметров СТ за счёт смещения при скрещивании значительно различающихся хромосом, которые переносятся на следующую итерацию генетического алгоритма.

При выполнении процедуры скрещивания $Crossover(\chi_{p1}, \chi_{p2})$ выполняется двухточечный кроссинговер хромосом родительских особей χ_{p1} , χ_{p2} , заключающийся в случайном выборе двух точек разрыва r_1 и r_2 хромосом.

При этом хромосома дочерней особи χ_c определяется следующим образом:

$$\chi_c = \langle \chi_{c,i} \rangle_{i=1}^{n_\chi}, \text{ где } \chi_{c,i} = \begin{cases} \chi_{p1,i}, & \text{если } i \in [r_1, r_2] \\ \chi_{p2,i}, & \text{если } i \notin [r_1, r_2] \end{cases}. \quad (4)$$

Мутация особей выполняется процедурой $Mutation(\chi)$ методом сальтации, заключающимся в выборе в хромосоме особи χ границ $j_0, j_1 \in [1, n_\chi - 1]$, где $j_0 < j_1$, в пределах которых производится замена значений бит хромосомы на противоположные. В результате формируется изменённая хромосома:

$$\dot{\chi} = \langle \dot{\chi}_i \rangle_{i=1}^{n_\chi}, \text{ где } \dot{\chi}_i = \begin{cases} \chi_i, & \text{если } i \in [j_0, j_1] \\ \chi_i \oplus 1, & \text{если } i \notin [j_0, j_1] \end{cases}. \quad (5)$$

Мутация особей выполняется лишь для особей, сгенерированных на текущем шаге работы генетического алгоритма, причём вероятность мутации p_{mut} определяется в соответствии с расстоянием Хэмминга между хромосомами её родительских особей следующим образом:

$$p_{mut} = 1 - \frac{D(\dot{\chi}_{p1}, \dot{\chi}_{p2})}{|\dot{\chi}_{p1}|}. \quad (6)$$

Данный механизм позволяет избежать сходимости популяции к локальным экстремумам критерия оптимальности решения за счёт высокой вероятности мутации для особей, имеющих слабо отличающиеся родительские хромосомы, и сохранить при этом наилучшие решения за счёт отсутствия мутации особей, переходящих из поколения в поколение.

На очередную итерацию генетического алгоритма переносятся лучшие особи текущей итерации, соответствующие наибольшим значениям $\omega(\psi)$, и их потомки в пропорции один к двум.

Критерием остановки генетического алгоритма является истечение заданного пользователем временного интервала.

Результатом работы генетического алгоритма является база данных, содержащая множество точек $\rho = \{\rho_i \mid \rho_i \in \Psi \times \Omega\}$, где $\rho_i = \langle \psi_i, \omega_i \rangle$, $\psi_i \in \Psi$, $\omega_i \in \Omega$, отражающих соответствие параметров СТ, обрабатываемого ССЗИ, обеспечиваемой им доступности информации. Наихудшие обнаруженные значения параметров доступности информации позволяют определять способность тестируемого ССЗИ удовлетворять требованиям, предъявляемым к компонентам защищаемой компьютерной сети.

Оценка надёжности генетического алгоритма

Разработанный генетический алгоритм предназначен для решения задачи поиска множества экстремумов неизвестной функции $\omega(\psi)$. Поэтому под его надёжностью подразумевается способность к обнаружению данных экстремумов в течение заданного ограниченного интервала времени.

Исследование надёжности было выполнено в соответствии со схемой, представленной на рис. 2, где МГА — модуль, реализующий разработанный генетический алгоритм; МИТ — модуль имитации тестирования ССЗИ с использованием СТ, обладающего заданными параметрами $\psi \in \Psi$, тестовой функции ТФ $\omega(\psi)$, отражающей зависимость вектора параметров доступности информации от параметров СТ; МАРТ — модуль анализа результатов тестирования.

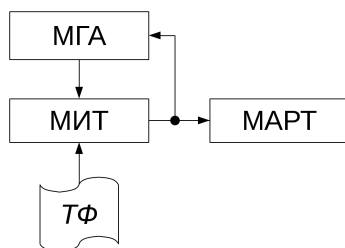


Рис. 2. Структура стенда для тестирования надёжности генетического алгоритма

Каждый из компонентов тестовой функции $\omega(\psi)$ содержит шумовую составляющую, затрудняющую поиск экстремумов, и информационную, содержащую искомые экстремумы. В качестве данных составляющих используются функции, широко применяемые при исследовании надёжности генетического алгоритма [3], критерием выбора которых является отсутствие ограничений по размерности вектора их аргументов для возможности представления многомерного пространства параметров. В качестве шумовой составляющей использована функция Растригина, особенностью которой является большое количество локальных экстремумов. В качестве информационной составляющей использована функция Михалевича, особенностью которой является наличие большого

количества локальных экстремумов и одного — глобального, которые занимают относительно небольшую часть области определения функции, что повышает сложность решения задачи поиска.

В результате выполнения серии из 100 запусков генетического алгоритма при выбранных случайным образом значениях коэффициентов тестовой функции были получены зависимости средних, минимальных и максимальных значений доли выявленных экстремумов. Графики, отражающие данные зависимости, приведены на рис. 3.

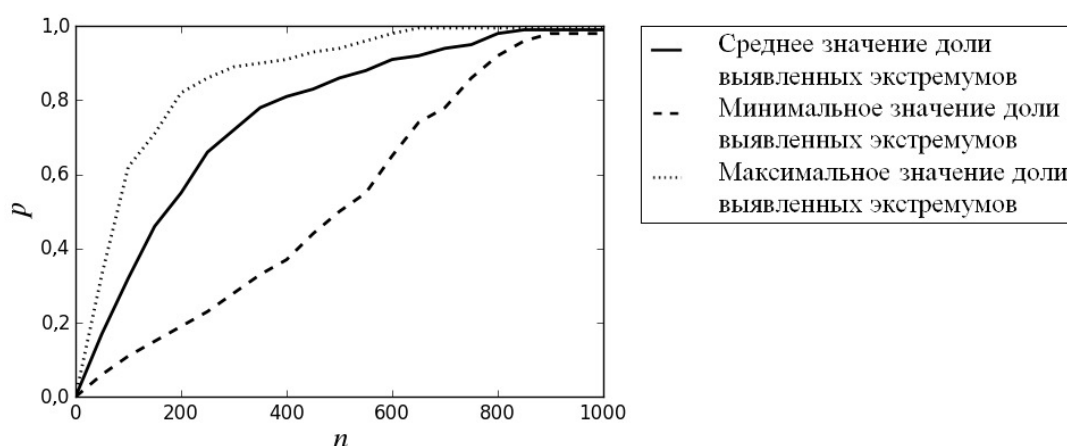


Рис. 3. Зависимость доли выявленных экстремумов от количества исследованных генетическим алгоритмом точек тестовой функции

Результаты эксперимента показали, что разработанный генетический алгоритм при количестве исследованных точек тестовой функции не менее 850 позволяет выявить не менее 98 % её локальных и глобальных экстремумов. В рамках методики IETF RFC 2544 [4], широко применяемой на сегодняшний день для тестирования различного оборудования, используемого в компьютерных сетях, устанавливается минимальная длительность генерации образца СТ, равная 120 с. При использовании указанной длительности исследование приведённого количества точек занимает менее 30 часов, что может служить подтверждением надёжности генетического алгоритма и его способности к решению задачи выявления экстремумов заданной функции в течение ограниченного интервала времени.

Разработанный генетический алгоритм позволяет автоматизировать процесс поиска уязвимостей ССЗИ к сетевым компьютерным атакам типа «отказ в обслуживании» путём подбора таких сочетаний параметров СТ атакующего воздействия, при которых ССЗИ оказывается неспособно обеспечить необходимый уровень доступности информации, обрабатываемой в компьютерной сети. При этом, в отличие от применяемых на сегодняшний день методик автоматизации тестирования (в частности, RFC 2544), предложенный генетический алгоритм позволяет решать данную задачу при большой размерности пространства поиска и позволяет учесть все известные параметры СТ, оказывающие влияние

на устойчивость ССЗИ к рассматриваемому типу атак.

Результаты практического применения генетического алгоритма

Для апробации предложенного генетического алгоритма разработан экспериментальный стенд, с помощью которого было произведено тестирование ряда образцов ССЗИ.

Результатом тестирования образца ССЗИ на данном стенде является множество критических областей пространства параметров сетевого трафика Ψ . Каждая из указанных областей определяет параметры сетевого трафика атаки типа «отказ в обслуживании», к которой уязвимо ССЗИ. Уязвимость ССЗИ выражается в том, что значение по крайней мере одного из параметров доступности информации $\omega \in \Omega$, обеспечиваемой им при обработке сетевого трафика атакующего воздействия, выходит за граничные значения, допустимые для данной компьютерной сети.

Критерием выявления критических областей выступило несоответствие требованиям стандарта МСЭ-Т Y.1541 [5] к сетям нулевого класса, предназначенным для исполнения приложений реального времени, которые накладывают следующие требования к доступности информации: относительная доля потерь пакетов q должна быть менее 0,001, джиттер $\sigma(\delta t_p)$ — менее 50 мс, а средняя задержка передачи пакетов $\bar{t}_d$ — менее 100 мс.

В частности, было выполнено тестирование фильтрующего маршрутизатора Cisco 2811, широко применяемого в сетях средних и крупных предприятий в качестве межсетевого экрана. В соответствии с документацией устройство имеет следующие технические характеристики:

- пропускная способность 61 Мбит/с;
- предельная интенсивность пересылки — 120000 пакетов/с;
- объём оперативной памяти 256 Мбайт.

Результаты тестирования маршрутизатора приведены на лепестковой диаграмме (рис. 4).

В результате тестирования были выявлены параметры сетевой компьютерной атаки типа «отказ в обслуживании», характеризующейся потерями пакетов $q \in [0,03, 0,06]$ и обусловленной достижением образцом ССЗИ предельного значения интенсивности передачи данных $I_d \in [61, 78]$ Мбит/с, соответствующего приведённому в документации образца.

Также была выявлена критическая область, соответствующая ранее неизвестной компьютерной атаке.

Данная критическая область показывает, что высокие значения потерь пакетов $q \in [0,05, 0,06]$ и джиттера $\sigma(\delta t_p) \in [51, 59]$ мс возникают при обработке образцом ССЗИ сетевого трафика относительно низкой интенсивности $I_d \in [51, 59]$ Мбит/с, где наблюдаются относительно высокие значения вариативности межпакетных временных интервалов $\sigma(t_p) \in [2,2, 2,9]$ мкс и средней интенсивности генерации потоков $\bar{n}_f \in [680, 815]$ потоков/с.

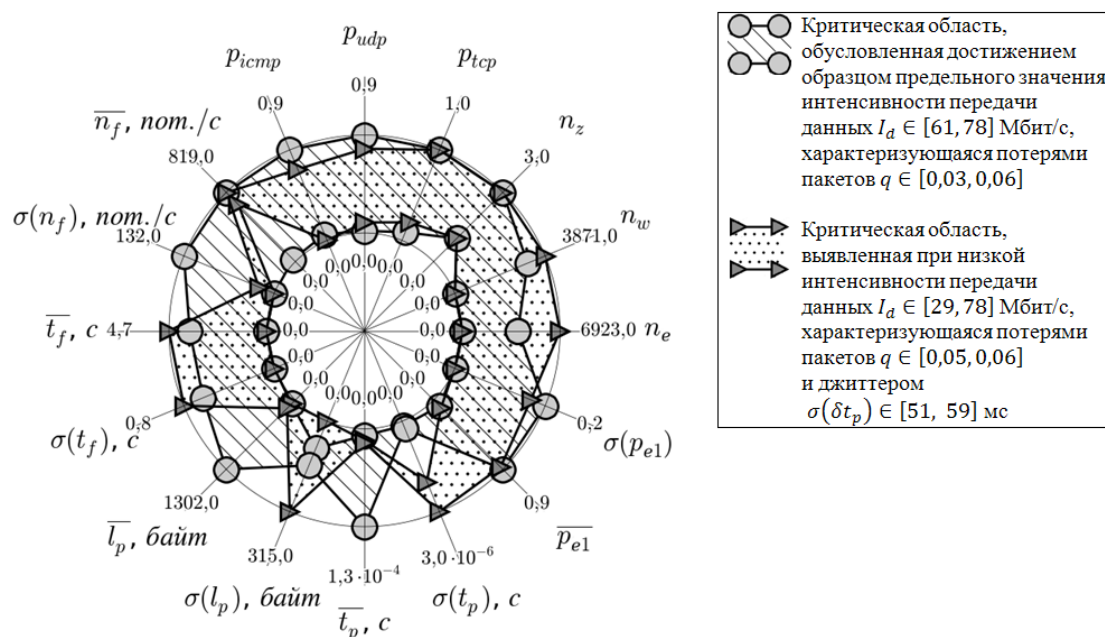


Рис. 4. Результаты тестирования фильтрующего маршрутизатора Cisco 2811

Выводы

Принцип работы разработанного генетического алгоритма тестирования ССЗИ основан на поиске экстремумов многомерной функции, отражающей зависимость между параметрами СТ, обрабатываемого ССЗИ, и обеспечиваемой им доступности информации. Данный поиск производится путём многократного тестирования ССЗИ с использованием различных образцов синтезированного СТ.

Результаты работы генетического алгоритма позволяют определить наихудшие показатели доступности информации, которые тестируемое ССЗИ способно обеспечить в процессе проведения сетевых компьютерных атак типа «отказ в обслуживании», что позволяет производить обоснованный выбор образцов ССЗИ, наиболее устойчивых к данному типу атак, и, таким образом, способствует обеспечению информационной безопасности информационно-телекоммуникационных систем и сетей.

ЛИТЕРАТУРА

1. Sridhar S. Denial of Service attacks and mitigation techniques: Real time implementation with detailed analysis. Essex : School of Computer Science and Electronic Engineering, University of Essex, 2011. 47 p.
2. Батищев Д.И., Неймарк Е.А., Старостин Н.В. Применение генетических алгоритмов к решению задач дискретной оптимизации. Нижний Новгород : изд-во ННГУ, 2007. 88 с.

3. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы; пер. с польск. И.Д. Рудинского. М. : Горячая линия–Телеком, 2006. 452 с.
4. McQuaid S., Bradner J. IETF RFC 2544: Benchmarking methodology for network interconnect devices. URL: <http://www.ietf.org/rfc/rfc2544.txt> (дата обращения: 25.11.2017).
5. ITU-T Y.1541: Network performance objectives for IP-based services. URL: <https://www.itu.int/rec/T-REC-Y.1541-201112-I> (дата обращения: 20.08.2017).

TESTING NETWORK SECURITY DEVICES AUTOMATION USING EVOLUTIONARY-GENETIC APPROACH

N.I. Sinadsky

Ph.D. (Eng.), Associate Professor, e-mail: nickis@e1.ru

A.V. Agafonov

Instructor, e-mail: avagaf@gmail.com

Institute of Radioelectronics and Information Technologies, Ural Federal University n.a.
first President of Russia B.N. Yeltsin, Yekaterinburg, Russia

Abstract. The article describes approach for automation of testing the immunity of network security devices against denial of service attacks using genetic algorithms.

Keywords: testing, network security devices, denial of service, genetic algorithm.

Дата поступления в редакцию: 20.04.2018

АНАЛИЗ ПРОБЛЕМ УПРАВЛЕНИЯ РАЗГРАНИЧЕНИЕМ ДОСТУПА В КРУПНОМАСШТАБНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Н.Ф. Богаченко

к.ф.-м.н., доцент, e-mail: nfbogachenko@mail.ru

Омский государственный университет им. Ф.М. Достоевского, Омск, Россия

Аннотация. В статье рассматриваются общие вопросы управления разграничением доступа к ресурсам крупномасштабных информационных систем (КМИС) с позиций формальных математических моделей. Анализируются свойства, присущие КМИС, и требования, предъявляемые к её политике безопасности, реализующей методы и правила разграничения доступа. Ставится задача разработки новых моделей, методов и алгоритмов управления разграничением доступа в КМИС.

Ключевые слова: разграничение доступа, политика безопасности, администрирование, автоматизация.

Введение

В настоящее время информационные системы и информационные технологии являются основой деятельности практически любого предприятия или организации. Согласно [37], информационные технологии — это «процессы, методы поиска, сбора, хранения, обработки, предоставления, распространения информации и способы осуществления таких процессов и методов». Информационная система определяется как «совокупность содержащейся в базах данных информации и обеспечивающих её обработку информационных технологий и технических средств» [37]. Таким образом, термин «информационная система» объединяет информационные ресурсы, программное обеспечение и компьютерное оборудование. В частности, крупные информационные системы, предназначенные для бизнеса, получили название «корпоративные информационные системы».

Принято различать три уровня функционирования объектов информатизации, обслуживаемых информационными системами:

- 1) микроуровень (учреждение, организация, предприятие);
- 2) мезоуровень (область, регион, отрасль);
- 3) макроуровень (государство, межгосударственная структура).

Наиболее востребованными являются информационные системы, обеспечивающие функционирование объектов 2-го и 3-го уровней. Как правило, это

масштабные, распределённые программно-технические комплексы с развитыми коммуникационными связями, обеспечивающие работу связанной группы организаций или предприятий. Такие системы принято называть *крупномасштабными информационными системами* (КМИС)¹.

Цель данной статьи — выявить основные проблемы, возникающие при разграничении доступа к информационным ресурсам КМИС, и определить пути их решения.

1. Крупномасштабные информационные системы и основы защиты информации

Можно выделить несколько причин появления и всё большего распространения КМИС.

1. Потребности бизнеса и государства во взаимодействии, в интеграции разрозненных информационных систем, обслуживающих эти структуры.

2. Потребности в расширении возможностей информационных систем управления ресурсами предприятия (Enterprise Resource Planning Systems, ERP-систем). Технологии развиваются в направлении выделения ядра ERP и создания дополнительных, например облачных, сервисов, специализированных приложений. Всё большее распространение приобретает мультивендорная стратегия, когда в одной информационной системе интегрируется программное обеспечение разных производителей.

3. Развитие концепции цифрового предприятия. Для цифрового предприятия важную роль играют удобные, безопасные, конкурентоспособные электронные ресурсы и информационные сервисы по работе с клиентом.

Использование КМИС позволяет вывести деятельность предприятий и организаций на качественно новый уровень производства и управления. Но в то же время их бизнес-процессы становятся чрезвычайно зависимыми от рисков и последствий сбоев применяемых информационных систем. Современные КМИС содержат тысячи или даже миллионы взаимодействующих компонентов аппаратного и программного обеспечения, в связи с чем возникает множество технических и организационных проблем — некоторые из них существовали ранее и теперь стали критическими, а другие возникли в последнее время в результате увеличения масштаба и степени взаимосвязи информационных подсистем.

Одним из основных источников возникающих проблем является уязвимость КМИС для вредоносных атак, направленных на реализацию угроз информационной безопасности [10]. Пробелы в безопасности КМИС могут приводить не только к нарушению конфиденциальности, потере, изменению или уничтожению информации, но и к сбою системы в целом, крупным финансовым и имиджевым потерям. Этому способствуют *масштабность, сложность и разнородность* крупномасштабных систем.

¹В англоязычных источниках имеется аналог этого понятия — Large-Scale Complex IT Systems [57,70].

1. По определению масштаб является отличительной чертой КМИС. Основные метрики измерения масштаба — это количество компонентов, содержащихся в системе, и количество пользователей, поддерживаемых системой.

2. Компоненты КМИС взаимодействуют и обмениваются информацией множеством способов, с многочисленными циклами обратной связи. Сбои в одной части системы могут непредсказуемым образом сказаться на работе КМИС в целом.

3. КМИС нередко создаются на основе объединения нескольких информационных подсистем. Этот процесс приводит к высокому уровню гетерогенности в системах и повышает потребность в функциональной совместимости между компонентами. При этом применение централизованного способа проектирования «сверху вниз» не всегда возможно. Часто КМИС представляет собой результат восходящей интеграции ряда отдельных подсистем.

Таким образом, дополнительные уязвимости КМИС связаны, в частности, с существенным ростом числа пользователей системы, наличием запутанных, плохо отслеживаемых информационных потоков, с распределённостью системы и необходимостью сопряжения подсистем с различными подходами к обеспечению информационной безопасности.

В свою очередь, одним из основных способов защиты информации является защита от несанкционированного доступа (НСД) [37]. Согласно [11], комплекс программно-технических средств и организационных мер по защите информации от НСД должен включать в себя подсистему управления доступом практически для всех классов информационных систем. При этом реализация дискреционного принципа контроля доступа относит используемые средства вычислительной техники ко 2-й группе защищённости, а реализация ещё и мандатной модели — к более защищённой 3-й группе [12]. В этом же документе поясняется, что «применение <...> средств криптографической защиты информации <...> может быть использовано для повышения гарантий качества защиты» [12]. Таким образом, наиболее распространённым и востребованным методом обеспечения конфиденциальности данных является управление доступом к ресурсам информационной системы. Стандарт ISO/IEC 27001/27002 также предусматривает в качестве основных мер защиты разграничение доступа пользователей к информационным ресурсам [27, 28].

С теоретической точки зрения классические и развивающиеся модели безопасности компьютерных систем строятся, исходя из понятия доступа субъекта к объекту, так как базируются на основной аксиоме компьютерной безопасности: «Все вопросы информационной безопасности в компьютерной системе определяются доступами субъектов к объектам»² [14]. Набор правил разграничения доступа в информационной системе, который принято называть *политикой разграничения доступа (политикой управления доступом)*, определяет основные принципы регулирования использования всех ресурсов системы.

Вышесказанное обосновывает утверждение о том, что задача адекват-

²Субъекты — это абстрактное представление активных сущностей системы (процессов, потоков, пользователей). Объекты — пассивные сущности системы (области памяти, файлы и т. д. — любые наборы данных, к которым можно обращаться как к целому).

ного построения политики разграничения доступа является не менее важной, чем, например, стойкость используемых криптографических алгоритмов. Программно-технические средства, реализующие политику разграничения доступа КМИС, составляют *подсистему разграничения доступа (подсистему управления доступом)*. Использование формальных моделей разграничения доступа на всех этапах жизненного цикла этой подсистемы обеспечивает успешное решение задачи выбора и обоснования механизмов реализации средств и методов защиты информационных ресурсов КМИС от несанкционированного доступа (см. рис. 1).



Рис. 1. Место формальных моделей разграничения доступа в системе комплексной защиты информации

2. Теоретические и практические подходы к разграничению доступа

Основы моделирования безопасного управления доступом и информационными потоками заложены в дискреционном (Discretionary Access Control, DAC) [47,51,52,60], мандатном (Mandatory Access Control, MAC) [46,62] и ролевом (Role-Based Access Control, RBAC) [49,50,67,68] принципах разграничения доступа. Классические формальные модели разграничения доступа получили широкое развитие и в работах российских исследователей [8,14,18,21,31,39] и др.

Вместе с тем практическая реализация классических моделей разграничения доступа в современных КМИС сталкивается с существенными трудностями [20]. Появляются, исследуются и находят практическое применение развивающиеся модели разграничения доступа. Развитие дискреционной модели связано с расширением понятия матрицы доступов: от типизированной (Typed Access Matrix, TAM) [66] через динамическую типизированную (Dynamic-Typed Access Matrix, DTAM) [71] до атрибутивной (Attribute-Based Access Matrix, ABAM) [76]. В работах [15–17, 19, 23] предлагается ДП-расширение дискреционной, мандатной и ролевой моделей. ДП-модели основаны на иерархии сущностей компьютерной системы. Данный подход позволяет проводить формальный анализ безопасности информационных систем с учётом инфор-

мационных потоков по памяти или по времени, динамических ограничений и возможности кооперации субъектов. Ролевой принцип управления доступом получил развитие в виде моделей логического разграничения доступа. В работе [26] приводятся результаты сравнительного анализа трёх современных моделей логического разграничения доступа: атрибутивной модели (Attribute-Based Access Control, ABAC) [38, 41, 53], сущностной модели (Entity-Based Access Control, EBAC) [1, 48] и реляционной модели на основе цепочек отношений, предложенной авторами для многопользовательских информационно-аналитических систем [6]. Управление доступом на основе задач (Task-Based Access Control, TBAC) — ещё одно направление развития ролевого подхода [26, 73]. Ряд работ посвящён оптимизации иерархических структур ролевой модели [5, 44, 55, 56, 58, 72, 74, 77]. Развитие получила так называемая проблема разработки ролей (Role Mining Problem, RMP), относящаяся к задачам интеллектуального анализа данных и использующая восходящий принцип проектирования политики ролевого разграничения доступа [2, 59, 61, 63, 69, 75]. Ведутся исследования по совмещению нескольких классических подходов в одной политике разграничения доступа [3, 4, 22, 24, 29, 45, 54, 64]. Наблюдается повышенный интерес к формальным моделям и методам разграничения доступа к информационным ресурсам в распределённых системах и облачных сервисах [13, 40, 42, 43, 65, 78].

Вместе с тем анализ современных работ, касающихся вопросов разграничения доступа, свидетельствует, что наряду с имеющимися существенными достижениями в области построения формальных моделей, методов и алгоритмов управления доступом к информационным ресурсам, методологические подходы к обеспечению безопасного управления доступом к объектам КМИС требуют дополнительных теоретических исследований с позиций масштабности, сложности и разнородности информационных систем.

В области реализации политик разграничения доступа активно ведутся программно-технические разработки. Системы управления доступом (СУД)³ представляют собой современные программно-технические комплексы, позволяющие решать ряд проблем, связанных с управлением доступом к информационным ресурсам КМИС. В качестве примеров таких систем можно указать следующие автоматизированные решения: СУД компании «Энкор» [32]; программное обеспечение по управлению доступом на основе криптографических алгоритмов компании «АВТОР» [34]; СУД компании «Куб», использующая специализированный документооборот заявок [33]; система управления идентификационными данными и доступом компании «Индид»; IBM Tivoli Identity Manager; Oracle Identity Management [35] и др.

Основная задача этих систем — автоматизация процессов управления пользователями, ролями, идентификационными данными, правами доступа. Преимущество автоматизированных решений заключается, в частности, в следующем:

³Более общие программные решения носят название «системы управления идентификационными данными и доступом пользователей» (Identity and Access Management System, IAMS) [36].

- 1) снижение вероятности ошибок при администрировании и, как следствие, снижение рисков информационной безопасности;
- 2) увеличение быстродействия за счёт сокращения времени предоставления доступа;
- 3) повышение производительности;
- 4) снижение накладных расходов.

Вместе с тем современные СУД обладают рядом недостатков. Во-первых, новые доступы регламентируются на основе предварительно настроенных политик и шаблонов. При появлении в системе нового пользователя к нему применяется один из заготовленных шаблонов пользовательских разрешений. Для КМИС доля шаблонов, задаваемых администратором безопасности «в ручном» режиме, будет существенной и потребует больших временных затрат. Объясняется это тем, что чем масштабнее объект, обслуживаемый информационной системой, тем более специфичны его бизнес-процессы, следовательно, более индивидуальна система разграничения доступа. Здесь возникает и вторая проблема — разрастание базы данных шаблонов, что требует дополнительных ресурсов.

Во-вторых, шаблоны создаются и корректируются, исходя из анализа бизнес-процессов. Тем самими используется нисходящая технология проектирования, что для крупномасштабных систем не всегда эффективно. При изменении бизнес-процессов, что неизбежно для КМИС с достаточно продолжительным жизненным циклом, отсутствие необходимых шаблонных разрешений может привести к утрате конфиденциальности или доступности информации. Одним из способов решения указанных проблем является построение адаптивной системы управления разграничением доступа [7], которая сможет автоматически подстраиваться под изменяющиеся параметры системы.

В-третьих, такое преимущество СУД, как интеграция с кадровыми сервисами компании, применительно к КМИС не всегда даёт положительный результат. По сути происходит отождествление управленческой иерархии компании с иерархией ролей политики разграничения доступа, которую предлагает СУД. Данный подход эффективен в ситуации, когда все подразделения в компании разные. Если же в иерархии имеются схожие подструктуры, то может происходить дублирование ролей или порождение ролей с «близкими» наборами полномочий. Это в свою очередь приводит к необходимости пополнения функционала СУД механизмами оптимизации иерархической структуры доступа.

Наконец, следует отметить, что основным подходом к разграничению доступа в СУД является ролевая модель. Вместе с тем на практике при построении подсистемы разграничения доступа КМИС нередки ситуации, требующие совместной реализации нескольких принципов разграничения доступа. К сожалению, вопросы автоматизации процессов управления доступом на основе мандатных и дискреционных подходов не получили должного развития в современных СУД.

Вышесказанное позволяет сделать вывод о том, что на сегодняшний день отсутствует единый подход к решению проблем, возникающих в процессе эксплуатации СУД. Очевидно, что современные технологии разграничения доступа к информационным ресурсам должны развиваться с учётом особенностей КМИС.

При этом разработка и реализация программно-технических решений, повышающих уровень безопасности и надёжности КМИС, должна быть поддержана дополнительными исследованиями фундаментальных вопросов информационной безопасности крупномасштабных систем, включающих в себя разработку новых моделей, методов и алгоритмов *автоматизации* процессов *управления разграничением доступа* к информации, то есть процессов построения и сопровождения (администрирования) политики разграничения доступа.

Для постановки основных задач, возникающих в процессе управления разграничением доступа в КМИС, необходимо:

- 1) выделить основные свойства (признаки) КМИС;
- 2) определить особенности разграничения доступа к информации в КМИС;
- 3) выявить существующие алгоритмические проблемы разработки политики разграничения доступа КМИС.

3. Особенности разграничения доступа к информации в крупномасштабных информационных системах

Растущее число проблем, возникающих при работе подсистемы разграничения доступа КМИС, объясняется отчасти тем, что вопросы управления доступом пользователей к информационным ресурсам находятся на стыке двух процессов: обеспечение безопасности и автоматизация [30]. При этом темпы разработки и внедрения КМИС настолько высоки, что детальной проработке функционала подсистемы, реализующей политику разграничения доступа, часто не уделяется должного внимания. Нередко обеспечение бизнес-процесса (то есть возможности получения доступа пользователя к приложению) получает больший приоритет, чем выполнение правил политики безопасности.

Вместе с тем существует ряд проблем, которые следуют из особенностей самих КМИС. Построение эффективной подсистемы разграничения доступа к информационным ресурсам КМИС требует особых подходов. Традиционные методики, методы и алгоритмы решения поставленной задачи ориентированы главным образом на обеспечение информационной безопасности объектов информатизации 1-го уровня. Чтобы определить специфику, присущую КМИС, требуется выделить и проанализировать основные свойства больших систем вообще, и КМИС в частности.

Общепринято считать, что определяющими свойствами большой системы (не обязательно информационной) являются масштабность, сложность и разнородность. Этот набор признаков следует детализировать. Согласно [9, 25], большие системы обладают следующими характерными свойствами:

- 1) наличие в системе подсистем, каждую из которых можно рассматривать как отдельную систему;
- 2) определение для каждой подсистемы иерархии целей и выполняемых функций;
- 3) выделение и декомпозиция процессов, связывающих подсистемы в единую структуру;

4) гетерогенность системы: мультизадачность элементов, различная природа внешних воздействий и выходных данных, многообразие ресурсов и информационных потоков;

5) системный подход к проектированию — необходима целостная концепция функционирования и развития предметной области, для которой создаётся система;

6) индивидуальность — большая система разрабатывается для вполне определённого объекта мезо- или макроуровня и содержит массу специфических черт;

7) продолжительность разработки и внедрения, как следствие — потребность в сопровождении.

Вышеперечисленные свойства необходимо конкретизировать применительно к КМИС. В работе [7] на примере государственных информационных систем приведено обоснование принадлежности КМИС к сложным адаптивным системам и выделены основные признаки КМИС.

1. Большой территориальный размах.
2. Сложное динамическое поведение.
3. Наличие разнородных, сложно взаимодействующих информационных элементов (узлов).
4. Многоуровневая иерархическая структура.
5. Многоцелевой характер управления и функционирования.
6. Наличие коллектива людей, осуществляющего управление.
7. Вероятностное внешнее воздействие.
8. Продолжительный жизненный цикл (время создания, период функционирования и эволюции).

Наиболее значимыми с точки зрения влияния на подсистему безопасности КМИС являются первые 4 свойства.

Указанные признаки обуславливают особенности в управлении доступом к информационным ресурсам КМИС. В [36] приведён перечень основных характеристик подсистемы безопасности КМИС: «Большое количество пользователей; разнообразие прикладных программ; разрозненные механизмы и политики доступа для различных ресурсов; большой объём обрабатываемой информации; отсутствие механизмов централизованного управления учётными записями; неконтролируемая деятельность привилегированных пользователей; необходимость соблюдения нормативных требований и обеспечения аудита». Для подсистемы разграничения доступа КМИС определяющими будут являться следующие особенности.

1. Существенное увеличение числа пользователей и объектов.

Информационные системы подразделяются на классы в зависимости от сложности решаемых ими задач и от числа поддерживаемых ими пользователей. Именно число пользователей, для которых данная система доступна, определяет её масштаб. Число пользователей и ресурсов КМИС в несколько раз превосходит эти показатели для информационных систем микроуровня. При построении подсистемы разграничения доступа необходимо учесть максимальное количество вариантов доступа. Следовательно, в КМИС число ос-

новных элементов подсистемы разграничения доступа таких, как роли, метки безопасности и т. д., также существенно возрастает.

2. Изменяемость политики разграничения доступа.

В силу сложного динамического поведения КМИС её подсистема разграничения доступа также должна приспосабливаться к изменяемым параметрам объектов и внешней среды.

3. Необходимость совмещения политик разграничения доступа.

Нередко КМИС разрабатывается не «с нуля», а строится из уже существующих отдельных компонент, часто разных производителей с разным подходом к разграничению доступа. Ситуация осложняется тем, что многие производственные процессы являются непрерывными, их нельзя приостановить для существенной модернизации или замены используемого программного обеспечения. Таким образом, подсистема разграничения доступа КМИС должна унаследовать принципы и методы разграничения доступа отдельных компонент, включённых в единую информационную систему.

4. Распределённость и иерархичность объектов доступа.

Большинство КМИС распределены в пространстве. При этом элементы (узлы) КМИС могут объединяться в сложную многоуровневую иерархическую структуру.

Перечисленные характеристики подсистемы разграничения доступа КМИС позволяют выделить и формализовать основные проблемы, возникающие в процессе управления разграничением доступа.

4. Проблемы управления разграничением доступа в крупномасштабных информационных системах

Большое число пользователей и ресурсов, изменяемость состава пользователей КМИС, территориальная распределённость приводят к возникновению множества подзадач администрирования, а, значит, повышается риск возникновения ошибок. Примерами могут являться неверное или несвоевременное назначение / лишение прав доступа, неправомерное использование идентификационных данных, нарушение правил политики безопасности и т. д. При этом автоматизация процессов распределения прав и полномочий пользователей чаще всего ограничивается установками «по умолчанию». Дальнейшее управление разграничением доступа проводится администратором безопасности «в ручном» режиме. Такой подход имеет ряд недостатков. В частности, злоумышленник может знать схему «по умолчанию», а сама схема требует большого объёма доопределений и индивидуальных настроек. Это приводит к востребованности алгоритмов, позволяющих частично или полностью автоматизировать процессы установки и администрирования прав доступа.

В связи с существенным увеличением числа пользователей и объектов КМИС и изменяемостью её политики разграничения доступа традиционный нисходящий подход к проектированию политики разграничения доступа сталкивается с трудностями на этапе реализации. Получает распространение восходящий принцип построения политики. В силу большого объёма информации,

необходимой для анализа КМИС, этот подход не может быть реализован в ручном режиме. Поэтому актуальной является задача разработки алгоритмов автоматизации процессов формирования политики разграничения доступа на основе подхода «снизу вверх».

Основная проблема, обусловленная изменчивостью политики разграничения доступа КМИС, — достаточно спонтанная эволюция политики, которая приводит к запутыванию существующих информационных потоков между сущностями системы, появлению дублирующих сущностей и т. п. Следствием этого является рост рисков утечки прав доступа. Подобные проблемы могут решаться за счёт оптимизации основных структур политики разграничения доступа в соответствии с новыми требованиями. При этом оптимальность должна пониматься в широком смысле как характеристика наиболее эффективной работы системы. Очевидно, что задача такой оптимизации многокритериальна, часть критериев могут противоречить друг другу. К сожалению, в большинстве современных информационных систем при реализации той или иной модели разграничения доступа вопросам оптимальности уделяется внимание лишь на этапе проектирования. Тогда как перестройка (оптимизация) политики разграничения доступа должна представлять собой итеративный процесс, основывающийся на непрерывном формировании новых и уточнении существующих требований предметной области.

В ситуации, когда КМИС создаётся на основе объединения нескольких информационных систем, каждая — со своей политикой разграничения доступа, — ставится задача совмещения действующих политик при условии безостановочного режима работы всех подсистем. В этом случае остро стоит вопрос эффективности процесса совмещения как с точки зрения информационной безопасности, так и с позиции затрат времени и человеческих ресурсов. Традиционный подход основан на поиске идеального решения, удовлетворяющего требованиям всех совмещаемых политик разграничения доступа. Но для крупномасштабных систем такое идеальное решение чаще всего не существует. Актуальным представляется использование алгоритмов поддержки принятия решений, позволяющих учесть ограничения совмещаемых политик разграничения доступа.

Кроме того, совмещение подходов к разграничению доступа приводит к ряду проблем, связанных с неоптимальностью результирующей модели разграничения доступа: увеличение времени на принятие решения о предоставлении или запрете доступа, появление противоречивых правил разграничения доступа. Последнее может приводить к отклонениям от политики разграничения доступа и, как следствие, к повышению рисков утечки прав и полномочий. В связи с этим по-прежнему актуальной остаётся указанная ранее задача разработки алгоритмов оптимизации информационных структур и правил разграничения доступа в подсистеме безопасности КМИС.

Иерархичность и распределённость объектов доступа КМИС ведёт к востребованности эффективных алгоритмов разграничения доступа к информационным ресурсам в распределённых системах и системах с многоуровневой иерархической структурой. При этом необходимо учитывать, что в распределённой

системе (облаке), как правило, отсутствует единая подсистема безопасности, и разделение доступа достигается с помощью средств шифрования и технологий распределения ключей доступа.

Приведённый перечень актуальных проблем, возникающих при построении и сопровождении политики разграничения доступа КМИС, определяет необходимость разработки новых научно-обоснованных моделей, методов и алгоритмов управления разграничением доступа в условиях обработки больших массивов данных о пользователях и информационных ресурсах. Для достижения поставленной цели требуется решение следующих основных задач (см. рис. 2).

1. Анализ и формализация классических моделей разграничения доступа с учётом особенностей КМИС.

2. Разработка методов и средств автоматизации процессов построения ролевого, мандатного и дискреционного разграничения доступа к объектам КМИС.

3. Разработка методов и средств оптимизации политики разграничения доступа КМИС на основе различных критериев.

4. Разработка методов и средств автоматизации процессов совмещения различных политик разграничения доступа КМИС.

5. Разработка методов и средств построения распределённой политики разграничения доступа КМИС.



Рис. 2. Связь основных задач с характерными особенностями разграничения доступа (РД) в КМИС

Основное требование к разрабатываемым подходам, моделям, методам, алгоритмам — автоматизация соответствующих процессов построения и сопровождения политики разграничения доступа. Так как полная автоматизация указанных процессов невозможна, в силу рассмотренных ранее свойств крупномасштабных систем, для решения поставленных задач целесообразно использовать подходы, основанные на процессах поддержки принятия решений. В частности, могут быть задействованы метод анализа иерархий и технологии интеллекту-

ального анализа данных (Data Mining). Новые подходы к управлению разграничением доступа должны приводить к построению масштабируемых алгоритмов, способных поддерживать работоспособность по мере увеличения числа пользователей.

Формализация процесса создания любой сложной системы представляет собой циклический (итерационный) процесс. Каждая итерация этого процесса состоит из следующих основных элементов: системный анализ, проектирование, реализация, сопровождение. Именно на этапе сопровождения могут быть сформулированы дополнительные требования к системе, направленные на её оптимизацию, получение нового или добавочного качества, повышение гибкости и способности к адаптации. Жизненный цикл политики разграничения доступа КМИС также должен включать в себя как стадию разработки (engineering) — постановка задачи, системный анализ, проектирование и реализация, так и стадию реконструкции (re-engineering) — администрирование, анализ и оптимизация. Следует отметить, что сформулированные задачи исследований в области управления разграничением доступа к ресурсам КМИС охватывают все этапы жизненного цикла политики разграничения доступа.

Заключение

Технологии управления доступом к информационным ресурсам должны основываться на формальных моделях и методах разграничения доступа, учитывающих масштабность, сложность и разнородность современных информационных систем.

На современном уровне развития СУД администрирование политики разграничения доступа характеризуется тем, что большинство операций по управлению разграничением доступа осуществляется администратором безопасности «в ручном режиме» или задаётся «по умолчанию» и через шаблонные решения. Это приводит к появлению ряда проблем при работе с КМИС.

Выявленные проблемы в построении и сопровождении политики разграничения доступа КМИС актуализируют разработку новых моделей, методов и алгоритмов, позволяющих частично или полностью автоматизировать процессы управления разграничением доступа в КМИС на всех этапах жизненного цикла политики разграничения доступа.

Автоматизация процессов управления разграничением доступа пользователей к ресурсам позволит повысить быстродействие и производительность СУД и снизить риски информационной безопасности за счёт снижения вероятности ошибок администрирования. При этом возможность автоматизации будет заложена в сами подходы к построению политики разграничения доступа, что позволит разрабатывать более надёжные и прогнозируемые СУД.

ЛИТЕРАТУРА

1. Афонин С.А. Система логического разграничения доступа для облачных информационных систем // Научный сервис в сети Интернет: труды XVIII Всероссий-

- ской научной конференции (19-24 сентября 2016 г., Новороссийск). М. : ИПМ им. М.В. Келдыша, 2016. С. 51–57. URL: <http://keldysh.ru/abrau/2016/17.pdf> (дата обращения: 23.05.2018).
2. Белим С.В., Богаченко Н.Ф. Использование решётки формальных понятий для построения ролевой политики разграничения доступа // Информатика и системы управления. 2018. № 1(55). С. 16–28.
 3. Белим С.В., Богаченко Н.Ф., Ракицкий Ю.С. Совмещение политик безопасности, основанное на алгоритмах поддержки принятия решений // Информационно-управляющие системы. 2016. № 5. С. 66–72.
 4. Белим С.В., Богаченко Н.Ф., Ракицкий Ю.С. Теоретико-графовый подход к проблеме совмещения ролевой и мандатной политик безопасности // Проблемы информационной безопасности. Компьютерные системы. 2010. № 2. С. 9–17.
 5. Богаченко Н.Ф. Локальная оптимизация политики ролевого разграничения доступа (Local Optimization of the Role-Based Access Control Policy) // CEUR Workshop Proceedings. 2017. V. 1965. URL: <http://ceur-ws.org/Vol-1965/paper14.pdf> (дата обращения: 23.05.2018).
 6. Васенин В.А., Иткес А.А., Шапченко К.А., Бухонов В.Ю. Реляционная модель логического разграничения доступа на основе цепочек отношений // Программная инженерия. 2015. № 9. С. 11–19.
 7. Володина А.А., Лёвкин И.М. Адаптивный подход к защите информации в больших информационных системах // Проблема комплексного обеспечения информационной безопасности и совершенствование образовательных технологий подготовки специалистов силовых структур : Межвузовский сборник трудов VI Всероссийской научно-технической конференции КОНФИБ'15. СПб. : Университет ИТМО, 2016. С. 65–73.
 8. Гайдамакин Н.А. Разграничение доступа к информации в компьютерных системах. Екатеринбург : Изд-во Урал. ун-та, 2003. 328 с.
 9. Горбачев В.Г. Некоторые проблемы создания больших информационных систем в учреждении // Центр системных исследований. Методология. 2009. URL: http://www.inmeta.ru/metod/big_syst.htm (дата обращения: 23.05.2018).
 10. Горбачевская Е.Н. Исследование механизмов защиты данных в корпоративных информационных системах // Вестник Волжского университета им. В.Н. Татищева. 2012. № 4(20). С. 18–23.
 11. Гостехкомиссия России. Руководящий документ «Автоматизированные системы. Защита от несанкционированного доступа к информации Классификация автоматизированных систем и требования по защите информации». Москва, 1992. 29 с.
 12. Гостехкомиссия России. Руководящий документ «Средства вычислительной техники. Защита от несанкционированного доступа к информации Показатели защищенности от несанкционированного доступа к информации». Москва, 1992. 21 с.
 13. Грушо А.А., Грушо Н.А., Тимонина Е.Е., Шоргин С.Я. Безопасные архитектуры распределённых систем // Системы и средства информатики. 2014. Т. 24, № 3. С. 18–31.
 14. Девянин П.Н. Модели безопасности компьютерных систем. Управление доступом и информационными потоками. 2-е изд., испр. и доп. М. : Научно-техническое издательство «Горячая линия — Телеком», 2013. 338 с.
 15. Девянин П.Н. О проблеме представления формальной модели политики безопасности операционных систем // Труды ИСП РАН. 2017. Том 29, Вып. 3. С. 7–16.

16. Девянин П.Н. О результатах формирования иерархического представления МРО-СЛ ДП-модели // Прикладная дискретная математика. Приложение. 2016. № 9. С. 83–87.
17. Девянин П.Н. Обзор семейства ДП-моделей безопасности логического управления доступом и информационными потоками в компьютерных системах // Информационные технологии. 2010. № 5. С. 20–25.
18. Девянин П.Н. Обзорные лекции по моделям безопасности компьютерных систем // Прикладная дискретная математика. 2009. Приложение № 2. С. 151–190.
19. Девянин П.Н. Ролевая ДП-модель управления доступом и информационными потоками в операционных системах семейства Linux // Прикладная дискретная математика. 2012. № 1. С. 69–90.
20. Жуковский О.И., Гриценко Ю.Б. Особенности создания системы информационной безопасности веб-ГИС ведения электронного генерального плана инженерной инфраструктуры // Доклады ТУСУР. 2013. № 2(28). С. 101–106.
21. Зегжда Д.П. Общая схема мандатных моделей безопасности и её применение для доказательства безопасности систем обработки информации // Проблемы информационной безопасности. Компьютерные системы. 2000. № 2. С. 28–32.
22. Иткес А.А. Объединение моделей логического разграничения доступа для сложно-организованных распределённых информационных систем // Проблемы информатики. 2010. № 1, С. 85–94.
23. Колегов Д.Н. Применение ДП-моделей для анализа защищённости сетей // Прикладная дискретная математика. 2008. № 1(1). С. 71–87.
24. Королев И.Д., Поддубный М.И., Носенко С.В. Применение сегмента матрицы доступов ХРУ в анализе информационной безопасности систем, реализующих мандатное разграничение доступа // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 101. С. 1811–1823.
25. Кротов А.А., Лупян Е.А. Обзор методов реструктуризации и интеграции информационных систем. М. : ИКИ РАН, 2000. URL: http://smis.iki.rssi.ru/students/alekro/Dissertation/Papers/Reengineering/my_review.html (дата обращения: 23.05.2018).
26. Лапин С.А. Модель разграничения доступа для систем, содержащих равнозначные объекты // Безопасность информационных технологий. 2016. № 2. С. 49–54.
27. Международный стандарт ISO/IEC 27001. URL: [http://www.pqm-online.com/assets/files/pubs/translations/std/iso-mek-27001-2013\(rus\).pdf](http://www.pqm-online.com/assets/files/pubs/translations/std/iso-mek-27001-2013(rus).pdf) (дата обращения: 23.05.2018).
28. Международный стандарт ISO/IEC 27002. URL: <http://pqm-online.com/assets/files/pubs/translations/std/iso-mek-27002-2013.pdf> (дата обращения: 23.05.2018).
29. Оленников А.А., Оленников Е.А., Захаров А.А., Широких А.В., Варнавский В.В. Разработка модели управления доступом для типовой медицинской информационной системы // Программные продукты и системы. 2016. № 1. С. 166–169.
30. Савельев М. Разграничение доступа – долой маски! // Информационная безопасность. 2007. № 4. С. 64–65.
31. Сизоненко А.Б. Арифметико-логическое представление матрицы доступа в дискреционной модели разграничения доступа // Вестник Воронежского института МВД

- России. 2012. № 3. С. 201–206.
32. Система управления доступом. URL: <https://n-core.ru/projects/solutions/acs.html> (дата обращения: 23.05.2018).
 33. Система управления доступом. URL: <https://www.cube-system.ru/products/sistema-upravleniya-dostupom-rasshiryayet-vozmozhnosti-produkta-kub> (дата обращения: 23.05.2018).
 34. Системы управления доступом к информационным ресурсам. URL: <http://avtor.ua/resheniya/sistemy-upravleniya-dostupom.html> (дата обращения: 23.05.2018).
 35. Системы управления идентификационными данными и доступом. URL: <http://altirix.ru/index.php/products/sistemy-upravleniya-identifikatsionnymi-dannymi-i-dostupom> (дата обращения: 23.05.2018).
 36. Сова А. Управление идентификационными данными и доступом – основа системы информационной безопасности КИС // Информационная безопасность. 2009. № 3. С. 34–35.
 37. Федеральный закон Российской Федерации от 27.07.2006 г. № 149-ФЗ «Об информации, информационных технологиях и о защите информации». URL: <http://base.garant.ru/12148555/#friends> (дата обращения: 23.05.2018).
 38. Чернов Д.В. О моделях логического управления доступом на основе атрибутов // Прикладная дискретная математика. Приложение. 2012. № 5. С. 79–82.
 39. Щеглов А.Ю. Защита компьютерной информации от несанкционированного доступа. СПб. : Наука и техника, 2004. 384 с.
 40. Ястребов И.С. Управление доступом в распределённых системах // Автоматизация процессов управления. 2010. № 2. С. 48–53.
 41. Al-Kahtani M.A., Sandhu R. A Model for Attribute-Based User-Role Assignment // Proceedings of the Computer Security Applications Conference. 2002. P. 353–362.
 42. Anitha R., Mukherjee S. Metadata Driven Efficient CRE based Cipher Key Generation and Distribution in Cloud Security // International Journal of Security and Its Applications. 2014. V. 8, No. 3. P. 377–392.
 43. Belim S.V., Bogachenko N.F. Distribution of Cryptographic Keys in Systems with a Hierarchy of Objects // Automatic Control and Computer Sciences. 2016. V. 50, No. 8. P. 777–786. URL: <http://link.springer.com/article/10.3103/S0146411616080071> (дата обращения: 23.05.2018).
 44. Belim S., Bogachenko N., Ilushechkin E. An analysis of graphs that represent a role-based security policy hierarchy // Journal of Computer Security. 2015. V. 23, No. 5. P. 641–657. URL: <http://content.iospress.com/articles/journal-of-computer-security/jcs532> (дата обращения: 23.05.2018).
 45. Belim S.V., Bogachenko N.F., Kabanov A.N., Rakitskiy Yu.S. Using the Decision Support Algorithms Combining Different Security Policies // IEEE Dynamics of Systems, Mechanisms and Machines (Dynamics). 15–17 Nov. 2016. URL: <http://ieeexplore.ieee.org/document/7818976/> (дата обращения: 23.05.2018).
 46. Bell D.E., LaPadula L.J. Secure Computer Systems: Unified Exposition and Multics Interpretation. Bedford, Mass.: MITRE Corp., 1976. MTR-2997 Rev. 1.
 47. Bishop M. Applying the Take-Grant Protection Model. Technical Report. Dartmouth College Hanover, NH, USA, 1990. 26 p.

48. Bogaerts J., Decat M., Lagaisse B., Joosen W. Entity-based access control: supporting more expressive access control policies // Proceedings of the 31st Annual Computer Security Applications Conference. ACM, 2015. P. 291–300.
49. Ferraiolo D., Cugini J., Kuhn R. Role-based access control: Features and motivations // Proceedings of Annual Computer Security Applications Conference. IEEE Computer Society Press. 1995. P. 249–255.
50. Ferraiolo D.F., Kuhn D.R. Role-Based Access Controls // Proceedings of 15th National Computer Security Conference. Baltimore MD, 1992. P. 554–563.
51. Harrison M.A., Ruzzo W.L., Ullman J.D. On Protection in Operating Systems // Communications of the ACM. 1975. P. 14–25.
52. Harrison M.A., Ruzzo W.L., Ullman J.D. Protection in Operating Systems // Communications of the ACM. 1976. V. 19, No. 8. P. 461–471.
53. Hu V.C., Ferraiolo D., Kuhn R., Schnitzer A., Sandlin K., Miller R., Scarfone K. Guide to Attribute Based Access Control (ABAC) Definition and Considerations // NIST Special Publication 800-162, January 2014.
54. Kocatürk M.M., Gündema T.I. Fine-Grained Access Control System Combining MAC and RBAC Models for XML // Informatica. 2008. V. 19, Issue 4. P. 517–534.
55. Koch M., Mancini L.V., Parisi-Presicce F. A Graph-Based Formalism for RBAC // ACM Transactions on Information and System Security. 2002. No. 5(3). P. 332–365.
56. Koch M., Mancini L.V., Parisi-Presicce F. Graph-based specification of access control policies // Journal of Computer and System Sciences. 2005. No. 71(1). P. 1–33.
57. Large-Scale Complex IT Systems. Development, Operation and Management / Editors: R. Calinescu, D. Garlan (Eds.). 17th Monterey Workshop 2012. Oxford, UK, March 19–21, 2012.
58. Leitner M. Delta Analysis of Role-Based Access Control Models // Lecture Notes in Computer Science. 2013. V. 8111. P. 507–514.
59. Li N., Li T., Molloy I., Wang Q., Bertino E., Calo S., Lobo J. Role Mining for Engineering and Optimizing Role Based Access Control Systems. Purdue University, IBM T.J. Watson Research Center, 2007.
60. Lipton R.J., Snyder L. A linear time algorithm for deciding subject security // Journal of ACM (Addison-Wesley). 1977. No. 3. P. 455–464.
61. Lu H., Vaidya J., Atluri V. Optimal boolean matrix decomposition: Application to role engineering // Proceedings of International Conference on Data Engineering (ICDE). 2008. P. 297–306.
62. McLean J. The specification and modeling of computer security // Computer. 1990. V. 23, Issue 1. P. 9–16.
63. Molloy I., Li N., Li T., Mao Z., Wang Q., Lobo J. Evaluating Role Mining Algorithms // Proceedings of ACM Symposium on Access Control Models and Technologies (SACMAT). 2009.
64. Ribeiro C., Zuquete A., Ferreira P., Guedes P. SPL: An Access Control Language for Security Policies with Complex Constraints // Proceedings of the Network and Distributed System Security Symposium. San Diego, CA. 2001. URL: https://scholar.google.co.uk/citations?view_op=view_citation&hl=ru&user=3PHaUacAAAAJ&citation_for_view=3PHaUacAAAAJ:LPZeul_q3PIC (дата обращения: 23.05.2018).
65. Ruj S., Nayak A., Stojmenovic I. DACC: Distributed Access Control in Clouds // Proc.

- of IEEE 10th International Conference "Trust, Security and Privacy in Computing and Communications (TrustCom)". 2011. P. 91–98.
66. Sandhu R.S. The Typed Access Matrix Model // Proceedings of IEEE Symposium on Security and Privacy. Oakland, California, 1992. P. 122–136.
67. Sandhu R., Coyne E., Feinstein H., Youman C. Role Based Access Control: A multidimensional view // Proceedings of 10th Annual Computer Security Applications Conference. Orlando, 1994. P. 54–62.
68. Sandhu R.S., Coyne E.J., Feinstein H.L., Youman C.E. Role-Based Access Control Models // IEEE Computer, 1996. No. 29(2). P. 38–47.
69. Schlegelmilch J., Steffens U. Role mining with ORCA // Symposium on Access Control Models and Technologies (SACMAT). 2005.
70. Sommerville I., Cliff D., Calinescu R., Keen J., Kelly T., Kwiatkowska M., McDermid J., Paige R. Large-scale Complex IT Systems // Communications of the ACM. 2011. V. 55(7).
71. Soshi M. Safety Analysis of the Dynamic-Typed Access Matrix Model // Proceedings of the 6th European Symposium on Research in Computer Security. 2000. V. 1895. P. 106–121.
72. Tahir M.N. Hierarchies in Contextual Role-Based Access Control Model (C-RBAC) // International Journal of Computer Science and Security (IJCSS). 2008. No. 2(4). P. 28–42.
73. Thomas R.K., Sandhu R.S. Task-based authorization controls (TBAC: a family of models for active and enterprise-oriented authorization management // Proceedings of the IFIP TC11 WG11.3 Eleventh International Conference on Database Security XI: Status and Prospects. 1997.
74. Toahchoodee M., Ray I., McConnell R.M. Using Graph Theory to Represent a Spatio-Temporal Role-Based Access Control Model // International Journal of Next-Generation Computing. 2010. No. 1(2). P. 231–250.
75. Vaidya J., Atluri V., Guo Q. The role mining problem: Finding a minimal descriptive set of roles // Proceedings of Symposium on Access Control Models and Technologies (SACMAT). 2007. P. 175–184.
76. Zhang X., Li Y., Nalla D. ABAM: An Attribute-Based Access Matrix Model // Proceedings of the ACM Symposium on Applied Computing (SAC 2005). Santa Fe, New Mexico, March 13–17, 2005. P. 359–363.
77. Zhang D., Ramamohanarao K., Versteeg S., Zhang R. Graph Based Strategies to Role Engineering // CSIIRW '10 Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research. 2010. Article No. 25.
78. Zukarnain Z.A., Khalid R. Quantum Key Distribution Approach for Cloud Authentication: Enhance Tight Finite Key. International conference on Computer Science and Information Systems (ICSIS'2014) Oct 17-18, 2014 Dubai (UAE). P. 28–33.

THE ANALYSIS OF PROBLEMS OF ACCESS CONTROL ADMINISTRATION IN LARGE-SCALE INFORMATION SYSTEMS

N.F. Bogachenko

Ph.D. (Phys.-Math.), Associate Professor, e-mail: nfbogachenko@mail.ru

Dostoevsky Omsk State University, Omsk, Russia

Abstract. In the article general questions of access control management for resources of large-scale information systems (LSIS) from positions of formal mathematical models are considered. The properties inherent in LSISs and the requirements imposed on its security policy, which implements the methods and rules of access control, are analyzed. The task is to develop new models, methods and algorithms for managing access control in LSIS.

Keywords: access control, security policy, administration, automation.

Дата поступления в редакцию: 25.05.2018

МЕТОД СТАТИЧЕСКОГО АНАЛИЗА ИСХОДНОГО КОДА ПРИЛОЖЕНИЙ ОПЕРАЦИОННОЙ СИСТЕМЫ ANDROID НА НАЛИЧИЕ ВРЕДОНОСНОГО КОДА

А.Н. Мироненко

к.т.н., доцент, e-mail: mironim84@mail.ru

Факультет компьютерных наук, Омский государственный университет им. Ф.М.
Достоевского, Омск, Россия

Аннотация. Данная работа посвящена теме защиты мобильных устройств, функционирующих на базе операционной системы Android, от программного обеспечения с вредоносной функциональностью. В работе представлена краткая классификация вредоносных воздействий на мобильные устройства, а также методов анализа программного обеспечения, предположительно являющегося вредоносным. Кроме того, предлагается метод статического анализа, осуществляющий поиск в исходном код программ вредоносных вставок. В основе предлагаемого метода лежит использование математических метрик сложности.

Ключевые слова: вредоносный код, компьютерный вирус, статический анализ, метрики сложности, Android.

Введение

Тема защиты мобильных устройств от вирусов является актуальной. В комментариях к статье 273 УК РФ «Создание, использование и распространение вредоносных программ для ЭВМ» даётся определение понятия «компьютерный вирус». Этим термином мы будем пользоваться в данной работе. Компьютерный вирус — это программа, которая умеет воспроизводить себя в нескольких экземплярах, модифицировать (изменять) программу, к которой она присоединилась, и тем самым нарушать её нормальное функционирование [1]. Предположение об актуальности темы защиты мобильных устройств от вирусов делается, исходя из того, что в настоящее время мобильные телефоны есть практически у каждого человека. Как правило, в мобильном устройстве хранится огромное количество конфиденциальной информации, например номера телефонов, заметки, содержащие пароли, и т. п. Наиболее популярной платформой, по информации сайта NetMarketShare (см. рис. 1), является операционная система Android, почти 67 % всех устройств работают на базе данной операционной системы.

Согласно статье [2], базы данных современным антивирусом содержат более 10 миллионов образов вредоносных Android-приложений, однако большая

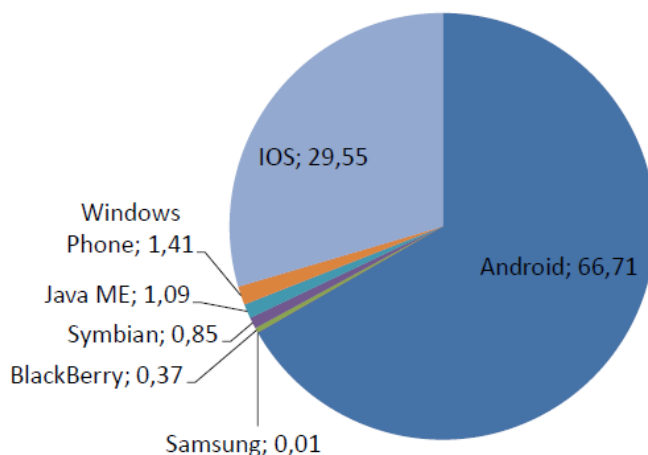


Рис. 1. Статистика мобильных операционных систем от NetMarketShare за февраль 2017

часть из них — это копии оригинальных вирусов. Проанализировав исходный код оригинальных вирусов, можно увидеть, что все они являются комбинацией небольшого количества уникальных функциональных блоков. Это объясняется тем, что вирусы для мобильных устройств, как правило, выполняют следующие однотипные задачи:

- 1) отправка СМС-сообщений на платные номера;
- 2) завладение конфиденциальной информацией пользователя (контакты, тексты сообщений, данные с карт памяти и т. п.);
- 3) сбор данных об устройстве;
- 4) получение прав администратора с целью установки программного обеспечения без ведома пользователя;
- 5) отслеживание и перехват логинов, паролей и данных банковских карт.

Всё вышесказанное, подтверждает актуальность защиты приложений операционной системы Android от вирусов.

1. Постановка задачи

В работе [3] приведена классификация методов анализа программного обеспечения, предположительно содержащего вредоносный код. Выделяют два типа анализа: динамический и статический. Динамические методы исследуют поведение программы, а именно, её взаимодействие с системой, собираемые данные, устанавливаемые сетевые соединения и т. п. Как правило, для анализа программы используют следующую информацию о ней и её поведении:

- 1) хеш-сумма APK-файла (MD5, SHA - 1 (256));
- 2) информация об отправляемых и получаемых по сети данных;
- 3) информация выполняемых чтении и записи файлов;
- 4) информация о службах, которые запускались;

- 5) информация о собранных и отправленных пользовательских данных;
- 6) информация о полученных приложением разрешениях;
- 7) информация об СМС-сообщениях, которые были отправлены, и осуществлённых вызовах.

Во время статического анализа программы исследуется её код. Основная задача такого анализа — это поиск части кода, которая осуществляет вредоносное воздействие. Идея применения математического аппарата для обнаружения вредоносного кода не нова. В работе [4] представлена новая теоретико-автоматная модель, способная стать основой для разработки антивирусов. Подход, описанный в работе, основан на технике проверки эквивалентности в алгебраических моделях последовательных программ. В работе [5] предложен метод поиска заимствований в программном коде с использованием метрик сложности. Алгоритм предлагаемого метода выглядит следующим образом:

- программа, подозреваемая в плагиате, представляется в виде точки $A(x_{a_1}, x_{a_2}, \dots, x_{a_n})$, где x_{a_i} — значение количественной метрики;
- программа, являющаяся оригинальной, так же представляется точкой $B(x_{b_1}, x_{b_2}, \dots, x_{b_n})$, где x_{b_i} — значение количественной метрики;
- размещаем точки A и B на n -мерном пространстве;
- считаем расстояние между точками A и B , если оно достаточно мало, то считается, что программа проверку не прошла. Кроме того, проведены результаты экспромта, подтверждающего возможность применения данного алгоритма.

В данной работе предлагается решать проблему поиска вредоносных вставок, взяв за основу идею, описанную и апробированную в работе [5].

2. Теория

Необходимо сформировать базу шаблонов вирусных вставок. Вычисляем метрики сложности от всех взятых, например из [2], фрагментов кода самых популярных типов вирусов. Тем самым получаем набор векторов, которые будут являться шаблонами для дальнейшего поиска вредоносной вставки в коде программы. Координаты отдельного вектора — это числовые значения метрик фрагмента кода конкретного типа вируса. Алгоритм поиска:

- берём N строк проверяемого кода и вычисляем его метрики сложности (точно такие же, как и тогда, когда мы формировали базу шаблонов вирусных вставок) — получаем вектор;
- сравниваем полученный вектор со всеми поочерёдно векторами из базы шаблонов вирусных вставок;
- если векторы совпали — вставка обнаружена, если нет — возвращаемся на «шаг 1»;
- увеличиваем N на 1 и повторяем «шаги 2–3».

Начальное значение N может быть равное 1, но лучше брать равное количеству строк кода самой «короткой» вставки из базы.

3. Выводы и заключение

В работе был предложен статический метод обнаружения вредоносных вставок в коде Android-программ. Предлагаемый метод основан на использовании метрик сложности программного кода. Возможность использования метрик сложности для сравнения двух программ была апробирована ранее в других работах, например [5], таким образом можно говорить, что и для решения поставленной задачи их применение также возможно. Тем не менее, необходима апробация данного алгоритма.

ЛИТЕРАТУРА

1. Ст. 272, «Уголовный кодекс Российской Федерации» от 13.06.1996 № 63-ФЗ (ред. от 19.02.2018). URL: <http://www.consultant.ru/cons/cgi/online.cgi?req=doc&base=LAW&n=291258&dst=976> (дата обращения: 27.04.2018).
2. Обзор фрагментов кода самых популярных типов вирусов. URL: <https://xakep.ru/2014/09/18/android-malware-source/> (дата обращения: 27.04.2018).
3. Скулкин О.В., Михайлов И.Ю., Макаров А.И. Принципы обнаружения вредоносных приложений для Android. URL: https://www.anti-malware.ru/analytics/Threats_Analysis/principles_detection_malicious_applications_Android#part2 (дата обращения: 27.04.2018).
4. Podlovchenko R.I., Kuzyurin N.N., Shcherbina V.S., Zakharov V.A. Using algebraic models of programs for detecting metamorphic malwares // *Fundamentalnaya i prikladnaya matematika*. 2009. V. 15, No. 5. P. 181–198.
5. МIRONENKO А.Н. Метод определения заимствований в программном коде с использованием его метрик сложности // Информационная безопасность и защита персональных данных. Проблемы и пути их решения: материалы VIII Всероссийской научно-практической конференции. 2016. С. 87–89.

THE METHOD OF STATIC ANALYSIS OF THE SOURCE CODE OF ANDROID APPLICATIONS FOR THE PRESENCE OF MALICIOUS CODE

A.N. Mironenko

Ph.D (Eng.), Associate Professor, e-mail: mironim84@mail.ru

Faculty of Computer Sciences, Dostoevsky Omsk State University, Omsk, Russia

Abstract. This work is devoted to the protection of mobile devices on the Android operating system from software with malicious functionality. The work presents a brief classification of malicious effects on mobile devices, as well as methods for analyzing software supposedly malicious. In addition, we offer a method of static analysis, which searches the source code for malware inserts. The proposed method is based on the use of mathematical complexity metrics.

Keywords: malicious code, computer virus, static analysis, complexity metrics, Android.

Дата поступления в редакцию: 03.05.2018

МЕТОД ПРИМЕНЕНИЯ (T, N) — ПОРОГОВОЙ СХЕМЫ В СТЕГАНОГРАФИИ

А.Н. Мироненко

к.т.н., доцент, e-mail: mironim84@mail.ru

Омский государственный университет им. Ф.М. Достоевского, Омск, Россия

Аннотация. Работа посвящена разработке метода скрытия информации в растровом изображении с использованием стеганографии совместно с криптографией. Предлагаемый метод позволяет решить проблему восстановления сообщения, если изображение, содержащее скрытые данные, было повреждено. Основная идея заключается в том, что данные с помощью стеганографии помещаются в изображение не целиком, а с использованием (t, n) — пороговой схемы, вставки каждой из частей происходит независимо. Разработано программное обеспечение для апробации предлагаемого метода. Проведена серия экспериментов, подтверждающих возможность применения предложенного метода.

Ключевые слова: схема разделения секрета, стеганография, LSB, схема Шамира.

Введение

Важным является вопрос сохранения и обеспечения конфиденциальности информации. С помощью специализированного программного обеспечения можно обеспечить надёжную передачу данных, зашифровав их. Однако в случае, когда злоумышленник не способен расшифровать данные, он может нарушить их целостность. Решением этой проблемы может быть использование стеганографии, т. е. сокрытие самого факта передачи секретной информации. Этой теме посвящено много научных работ, например [1–4]. В статье [5] рассмотрены основные тенденции в стеганографии.

В работе [6] описывается разработка защищённой системы передачи данных. Идея системы состоит в использовании на платформе Android двух криптографических алгоритмов: RSA и AES, совместно с LSB для реализации стеганографии. Объединение этих трёх алгоритмов позволяет создать защищённую систему связи на платформе Android.

В [7] предлагается стеганографическая система защиты информации на основе предлагаемого оригинального алгоритма для встраивания информации с перекрывающимися блоками изображений в строках и столбцах. Показано, что эта стеганографическая система сохраняет устойчивость к пассивным стеганоаналитическим атакам с перекрытием блоков до 24×24 пикселей, и при этом

значение перекрытия более стабильно, чем стандартный и улучшенный метод стеганографии, основанный на прямом распространении спектра.

В статье [8] автор предлагает стеганографический алгоритм, использующий пару ключей: открытый и закрытый, чтобы генерировать псевдослучайную последовательность, которая указывает, где будет храниться секретная информация. Перед вставкой сообщения изображение претерпевает несколько преобразований. Для извлечения информация преобразования применяются в обратном порядке. В стеганографии можно выделить несколько проблем:

- обнаружение стеганографической вставки (стегоанализ);
- надёжность сокрытия информации при её передаче.

Решению первой проблемы посвящена работа [9]. Решение второй проблемы предлагается в [10], в этой работе основное внимание уделяется интеграционным схемам, таким как OFDM, CDMA и MC-CDMA, со стеганографией и методами шифрования изображений для создания беспроводных систем со встроенной функцией обеспечения безопасности. В статье [11] авторы рассматривают метод унификации криптографии и стеганографии. Эта работа также посвящена проблеме повышения надёжности скрытия информации при её передаче, т. е. решению задачи восстановления информации в случае повреждения изображения, содержащего скрытые данные. Рассмотрим возможность комбинирования стеганографического метода и (t, n) -пороговой схемы. Кроме того, проведём анализ существующих алгоритмов стеганографии и пороговых схем.

1. Постановка проблемы и её решение

Проведём анализ существующих методов стеганографии и выберем наиболее подходящий. Рассмотрим алгоритмы для встраивания информации в часть исходного изображения. Преимуществом этих алгоритмов является то, что для внедрения не нужно выполнять сложные линейные преобразования изображения. Информация внедряется путём манипуляции с яркостью или цветовыми компонентами изображения.

1. Алгоритм Каттера. В этом алгоритме информация будет встраиваться в канал синего цвета, т. к. изменения этого цвета наименее заметны для человеческого глаза. В изображении выбирается псевдослучайная позиция, в которую будет вставлена информация и далее — в канал синего цвета. Информация вносится путём изменения яркости. В работе [12] рассматривается возможность использования стеганографического метода Kutter–Jordan–Bossen, позволяющего скрыть информацию о видеопоследовательности. Кроме того, в работе предлагается критерий выбора изображения контейнера.

2. Алгоритм Лангелаара. Этот алгоритм работает с блоками размером 8×8 пикселей от исходного изображения. Первоначально создаётся псевдослучайная маска размером 8×8 пикселей, состоящая из нулей и единиц. На первом этапе каждый блок делится на пару подблоков в зависимости от значения маски. Затем для каждого из них вычисляется среднее значение яркости. На втором этапе выбирается произвольный порог, затем биты встраиваются следующим образом: 1 записывается, если разница между средними значениями

яркости больше, чем выбранный порог; 0 записывается, если разница между средними значениями яркости меньше выбранного порога. Если это условие не выполняется, изменения будут происходить в значениях пикселей второго субблока.

3. Алгоритм Ронгена. Встраиваемая информация представляется в виде двумерной матрицы, состоящей в равных частях из единиц и нулей. На основе некоторой характеристической функции, вычисленной локально в процессе анализа соседних пикселей, определяются те пиксели, в которые может быть встроена информация. Количество этих пикселей составляет около 0,01 от общего числа, поэтому не все единицы будут внедрены в эти пиксели. Чтобы увеличить количество этих пикселей, первоначально предлагаем осуществлять незначительное искажение изображения.

4. Встраивание в наименее значащие биты (LSB). Суть этого метода заключается в том, что наименее значащие биты изображения несут в себе наименьшую информацию. Когда они заменяются на биты скрываемого сообщения, отличие получившегося изображения от исходного практически незаметно человеческому глазу. Этот метод позволяет встраивать большие объёмы информации, что является значительным плюсом, кроме того, он легко реализуем [13–15]. После анализа известных методов сокрытия данных в растровых изображениях и изучения работ [16, 17] для решения поставленной задачи был выбран метод LSB.

Проанализируем существующие пороговые схемы.

1. Пороговая схема Шамиром [18]. Пусть задано конечное поле G . Фиксируем n разных ненулевых нескретных элементов этого поля. Каждый из этих элементов присваивается определённому члену группы. Далее берём произвольное множество t элементов поля G , из них строим многочлен $f(x)$ степени $t - 1$, $1 < t \leq n$ над полем G . Получив многочлен, вычисляем его значение в нескретных точках и сообщаем результаты соответствующим членам группы. Чтобы восстановить секрет, можно использовать формулу интерполяции, например формулу Лагранжа.

2. Схема Блэкли. Секрет, который должны быть разделён в схеме Блэкли, является одной из координат точки в m -мерном пространстве. Доли секретов, распределённых между сторонами, являются уравнениями $(m - 1)$ -мерных гиперплоскостей. Чтобы восстановить точку, необходимо знать m уравнений гиперплоскостей.

3. Секретное разделение с использованием китайской теоремы об остатках (схема Миньотта, схема Асмута–Блума). Для некоторого числа (по схеме Миньотта это сам секрет, в схеме Асмута–Блума — некоторое производное число), рассчитываются остатки от деления на последовательность чисел, которые разделены между сторонами. Из-за ограничений на последовательность чисел только определённое количество сторон может восстановить секрет. Для дальнейшего анализа были выбраны две схемы: 1 и 3. Схема Блэкли не была включена в анализ, потому что она менее эффективна, чем схема Шамира: по схеме Шамира каждая доля такого же размера, как секрет, а по схеме Блэкли каждая доля в несколько раз больше. Результаты сравнения схем представлены

в таблицах 1 2. Была выбрана схема Шамира.

Таблица 1. Схема Шамира

Критерий сравнения	$t = 5, n = 3,$ $ M = 512$	$t = n = 32,$ $ M = 512$	$t = n = 128,$ $ M = 512$
Время разделения	5 мс	7 мс	75 мс
Время восстановления	1 мс	2 мс	27 мс
Объём требуемой памяти	1560 байт	6720 байт	100608 байт

Таблица 2. Схема Асмута–Блума

Критерий сравнения	$t = 5, n = 3,$ $ M = 512$	$t = n = 32,$ $ M = 512$	$t = n = 128,$ $ M = 512$
Время разделения	130 мс	145 мс	3032 мс
Время восстановления	1 мс	4 мс	97 мс
Объём требуемой памяти	3456 байт	274432 байт	4243456 байт

Идея предлагаемого в данной работе метода заключается в объединении стеганографического метода LSB с (t, n) -пороговой схемой, что позволяет при повреждении стегоконтейнера на $n - t$ частей, где $t \leq n$ и n — это общее количество долей секрета, а t — минимально необходимое количество долей секрета для его успешного восстановления, всё равно обеспечить надёжное восстановление секрета. Предлагаемый подход можно описать в два этапа:

- встраивание информации в изображение–контейнер, рис. 1;
- восстановление информации рис. 2.

Встраиваемая информация обрабатывается в соответствии с выбранной (t, n) -пороговой схемой, т. е. формируются доли секрета. Исходное изображение, в которое будет осуществляться вставка, делится на n частей (блоков) по ширине, в соответствии с (t, n) -пороговой схемой, используемой для формирования долей секрета. Каждая из полученных частей рассматривается как независимое изображение. Изображение представлено в виде массива бит, в котором для кодирования каждого цвета пикселя используется 8 бит. Секрет записывается внесением изменения в младшие разряды. В первый пиксель изображения пишем длину части секрета, а в последующие — саму часть секрета. Такие изменения произойдут со всеми n блоками. После завершения процедуры встраивания изображений контейнер формируется из n частей, которые уже содержат доли секрета. Чтобы извлечь информацию, необходимо знать количество частей, на которые было разделено исходное изображение, и

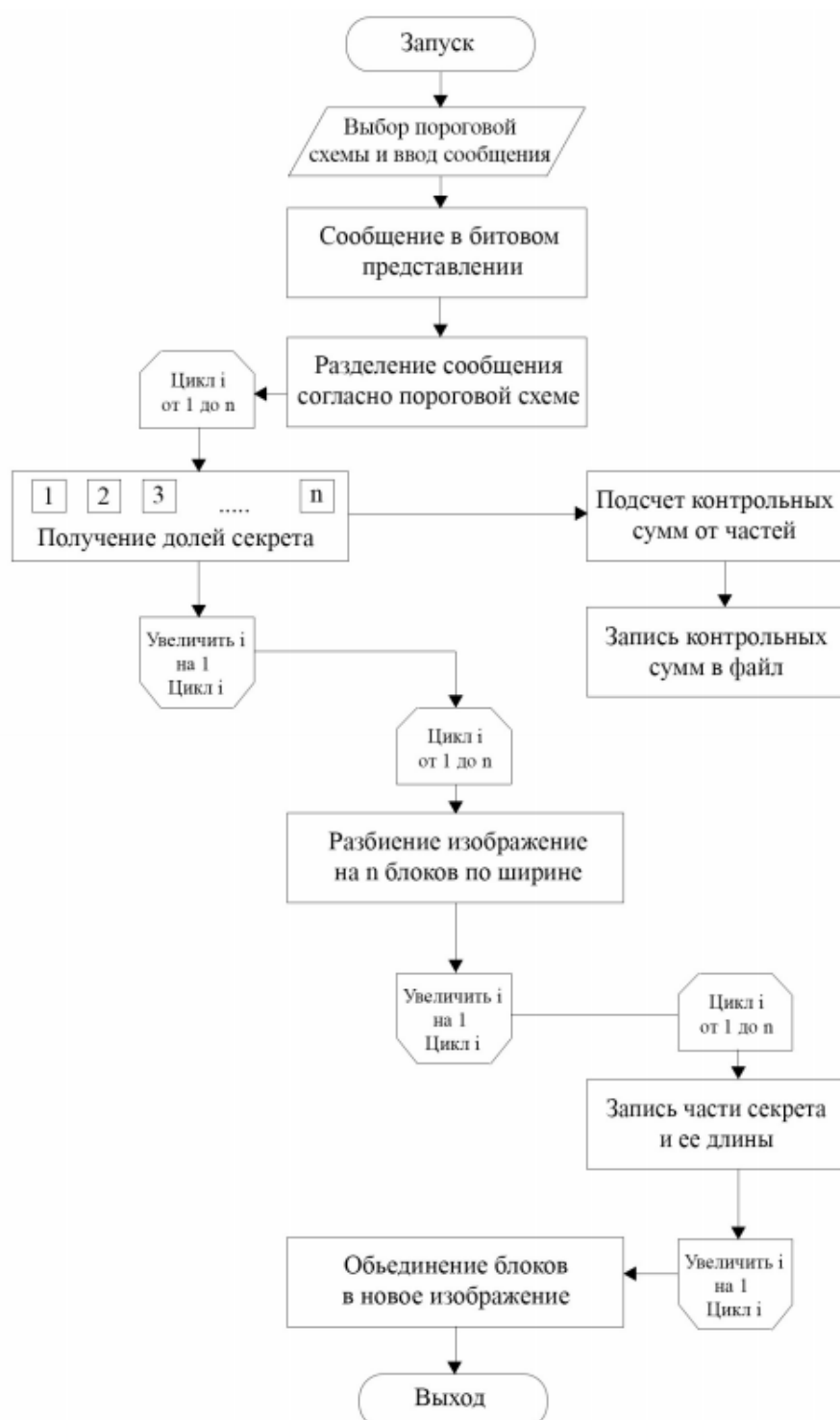


Рис. 1. Блок-схемы встраивания информации

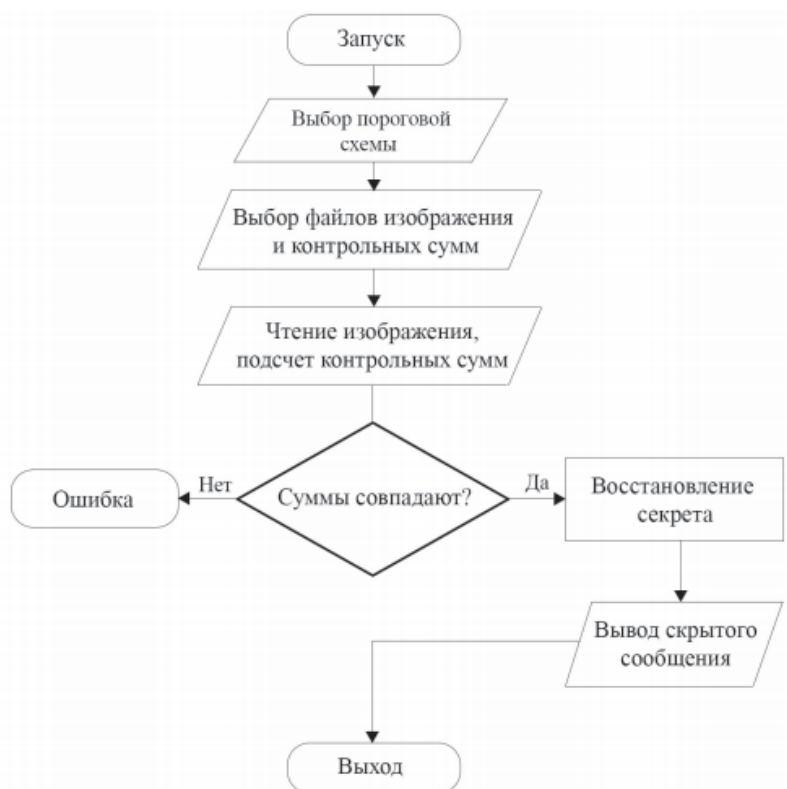


Рис. 2. Блок-схема работы алгоритма восстановления секрета

его длину. Затем из каждой части изображения-контейнера считывается первый пиксель и проверяется наличие записи длины секретной части скрытого сообщения. Если такая информация присутствует, то доля секрета считывается. Затем проверяются контрольные суммы. Они будут подсчитываться снова от каждой части изображения и будут сравниваться с теми, которые были записаны в файле. В случае их совпадения доля тайны будет записана в массив, который будет передан на вход функции восстановления секрета.

2. Апробация

Для апробации предложенного метода была реализована программа на языке C#. Блок-схемы работы программы представлены на рис. 3.

Был проведён ряд экспериментов. Для большей вероятности восстановления тайны лучше всего выбрать как можно меньше t и как можно больше n . На изображение-стегоконтейнер наносились следующие повреждения: наложение «шума», вырезание части картинки, наложение посторонних изображений. В случаях, когда изображение повреждалось путём вырезания его части или наложением посторонних изображений, эксперименты проходили удачно,

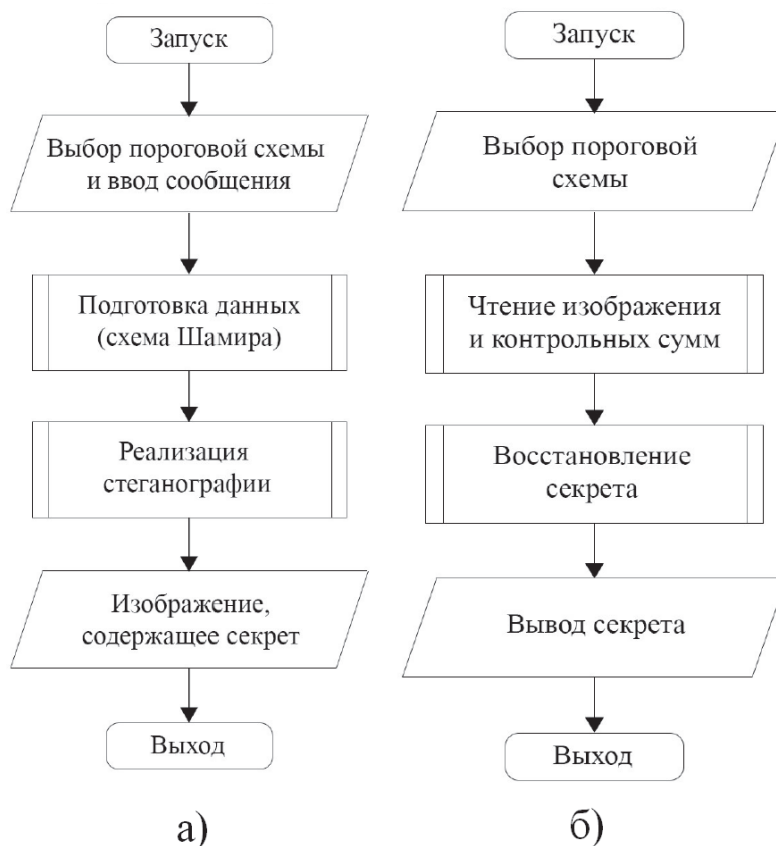


Рис. 3. Блок-схема работы программы: а) разделение секрета, б) восстановление секрета

т. е. изображение восстанавливалось при повреждении менее 60 % контейнера. Результат экспериментов, когда повреждение производилось путём нанесения «шума», представлен в таблице 3, где «+» — восстановить информацию удалось, «—» — восстановить информацию не удалось.

Таблица 3. Результаты эксперимента для различных значений t и n пороговой схемы

Уровень «шума» (%)	$t = 3,$ $n = 10$	$t = 6,$ $n = 10$	$t = 8,$ $n = 10$	$t = 10,$ $n = 10$
0,05	+	+	+	+
0,10	+	+	+	+
0,15 — 0,19	+	+	+	+
0,2	—	—	—	—

Значительно улучшить результаты можно путём выбора долей для восстановления информации по определённому алгоритму. Задача поиска алгоритма

выбора долей не ставилась в данной работе. Необходимо было проверить гипотезу о возможности объединения стеганографического метода LSB с (t, n) -пороговой схемой. Поиск алгоритма выбора долей может являться направлением последующих работ по данной теме.

3. Выводы

В работе был предложен метод объединения стеганографического метода LSB и (t, n) -пороговой схемы. Проведён анализ стеганографических методов и пороговых схем. Разработано программное обеспечение, проведена апробация. В ходе экспериментов было определено, что:

- в случае повреждения стегоконтейнера наложением постороннего изображения или путём удаления части изображения-стегоконтейнера порог успешного восстановления встраиваемой информации 60 %;
- в случае повреждения стегоконтейнера наложением «шума» порог успешного восстановления 0,19 %.

В целом можно говорить о возможности применения предлагаемого метода для повышения надёжности сокрытия информации при её передаче или порче стегоконтейнера.

ЛИТЕРАТУРА

1. Al-Afandy K., Faragallah O., Elmhawwy A. High security data hiding using image cropping and LSB least significant bit steganography // Information Science and Technology (CiSt). 4-th IEEE International Colloquium. 24–26 Oct., 2016. DOI: <https://doi.org/10.1109/CIST.2016.7805079>.
2. Chitradevi B., Thinaharan N., Vasanthi M. Data Hiding Using Least Significant Bit Steganography in Digital Images. 2017. DOI: <http://doi.org/10.5281/zenodo.262996>
3. Al-Dmour H., Al-Ani A., Nguyen H. An efficient steganography method for hiding patient confidential information // Conf Proc IEEE Eng Med Biol Soc. 2014. DOI: <https://doi.org/10.1109/EMBC.2014.6943569>.
4. Kaur S., Bansal S., Bansal R.K. Steganography and classification of image steganography techniques // Computing for Sustainable Global Development (INDIACom). International Conference. 5–7 March, 2014. DOI: <https://doi.org/10.1109/IndiaCom.2014.6828087>.
5. Zielinska E., Mazurczyk W., Szczypiorski K. Trends in Steganography // Communications of the ACM. 2014. No. 03, issue 57. P. 86–95. DOI: <http://dx.doi.org/10.1145/2566590.2566610>.
6. Kandul A., More A., Davalbhakta O., Artamwar R., Kulkarni D. Steganography with Cryptography in Android // Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014. Advances in Intelligent Systems and Computing. V. 328. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-12012-6_7.
7. Baltaev R.H., Lunegov I.V. Algoritm vstraivaniya i izvlecheniya informacii v nepodvizhnye cifrovye izobrazheniya stojkij k passivnym stegoanaliticheskim atakam

- // Voprosy bezopasnosti. 2016. No. 6. P. 24–35. DOI: <https://doi.org/10.7256/2409-7543.2016.6.21252>. Available at: http://e-notabene.ru/nb/article_21252.html (accessed 10 January 2018).
8. Soria-Lorente A., Berres S. A Secure Steganographic Algorithm Based on Frequency Domain for the Transmission of Hidden Information // Security and Communication Networks. 2017. Article ID 5397082. 14 p. DOI: <https://doi.org/10.1155/2017/5397082>.
 9. Belim S V., Vilkhovskiy D.E. Usage of analytic hierarchy process for steganographic inserts detection in images // Dynamics of Systems, Mechanisms and Machines (Dynamics). 2016. 15–17 Nov., 2016. DOI: <https://doi.org/10.1109/Dynamics.2016.7818977>.
 10. Praveenkumar P., Thenmozhi K., Rayappan J.B.B., Amirtharajan R. Inbuilt Image Encryption and Steganography Security Solutions for Wireless Systems: A Survey // Research Journal of Information Technology. 2017. No. 9. P. 46–63. DOI: <http://dx.doi.org/10.3923/rjit.2017.46.63>.
 11. Praveenkumar P., Thenmozhi K., Rayappan J.B.B., Amirtharajan R. Cryptic Cover for Covered Writing: A Pre-Layered Stego // Information Technology Journal. 2014. No. 13. P. 2524–2533. DOI: <http://dx.doi.org/10.3923/itj.2014.2524.2533>.
 12. Lysenko N., Labkov G. Applying of Kutter-Jordan-Bossen steganographic algorithm in video sequences // Young Researchers in Electrical and Electronic Engineering (EIConRus). IEEE Conference of Russian. 1-3 Feb., 2017. DOI: <https://doi.org/10.1109/EIConRus.2017.7910651>.
 13. Zeeshan M., Ullah S., Anayat S., Hussain R.G., Nasir N. A Review Study on Unique Way of Information Hiding: Steganography // International Journal on Data Science and Technology. 2017. V. 3, No. 5. P. 45–51. DOI: <http://dx.doi.org/10.11648/j.ijdst.20170305.11>.
 14. Thangadurai K., Sudha D.G. An analysis of LSB based image steganography techniques // Computer Communication and Informatics (ICCCI). International Conference. 3-5 Jan., 2014. DOI: <https://doi.org/10.1109/ICCCI.2014.6921751>.
 15. Liu J., Tian Y., Han T. et al. Stego key searching for LSB steganography on JPEG decompressed image // Sci. China Inf. Sci. 2016. No. 59. P. 32105. DOI: <https://doi.org/10.1007/s11432-015-5367-x>.
 16. Fkirin A, Attiya G., El-Sayed A. Steganography Literature Survey, Classification and Comparative Study // Communications on Applied Electronics. 2016. No. 5(10). P. 13-22. DOI: <https://doi.org/10.5120/cae2016652384>.
 17. Wiseman S.R. Defenders Guide to Steganography // Deep Secure Technical Report DS-2017-2. DOI: <https://doi.org/10.13140/rg.2.2.21608.98561>.
 18. Luo P, Yu-Lun L.A., Wang Z. Hardware Implementation of Secure Shamir's Secret Sharing Scheme // High-Assurance Systems Engineering (HASE). IEEE 15th International Symposium. 9–11 Jan., 2014. DOI: <https://doi.org/10.1109/HASE.2014.34>.

THE METHOD OF APPLYING (T, N) — THRESHOLD SCHEME IN STEGANOGRAPHY

A.N. Mironenko

Ph.D. (Eng.), Associate Professor, e-mail: mironim84@mail.ru

Faculty of Computer Sciences, Dostoevsky Omsk State University, Omsk, Russia

Abstract. The work is devoted to the development of the method of information concealment in a raster image using steganography together with cryptography. The suggested method allows to solve the problem of message recovery if the image containing the hidden data was damaged. The basic idea is that the data with the help of steganography is not placed in the image entirely, but using the (t, n) — threshold scheme, the insertion of each part occurs independently. A software has been developed to test the proposed method. A series of experiments confirming the possibility of applying the proposed method is carried out.

Keywords: Threshold scheme, steganography, LSB, Shamir's secret sharing.

Дата поступления в редакцию: 03.05.2018

Авторам

Предоставляемые данные и документы

Автор предоставляет в редакцию:

- рукопись статьи в формате $\text{\LaTeX}$ (см. требования к оформлению);
- список из трёх экспертов по тематике статьи, давших согласие написать рецензию на представленную работу¹;
- экспертное заключение о возможности открытого опубликования.

Лицензирование

Согласно ГК РФ ст. 1286 лицензионный договор с автором для публикации в периодических изданиях может быть заключён в устной форме. Сам факт получения рукописи статьи редколлегией журнала «Математические структуры и моделирование» является акцептом (принятием) лицензионного договора.

Все статьи в журнале «Математические структуры и моделирование» публикуются под лицензией Creative Commons Attribution 4.0 International (CC-BY). Текст лицензии находится по адресу <https://creativecommons.org/licenses/by/4.0/legalcode>.

Требования к оформлению рукописи

К публикации принимаются рукописи объёмом не более 16 страниц.

Авторам необходимо предоставить следующую информацию на русском и английском языках:

- название статьи;
- список авторов с указанием
 - фамилии, имени и отчества,
 - учёного звания,
 - учёной степени,
 - должности,
 - места работы или учёбы,
 - действующего адреса электронной почты;
- аннотация (абстракт) объёмом от 100 до 250 слов;
- список ключевых слов.

Автор также указывает УДК (универсальный десятичный код) статьи. Его можно подобрать по тематике статьи в справочнике <http://msm.univer.omsk.su/udc/>.

Библиографические ссылки оформляются согласно ГОСТ 7.0.5–2008.

Рукопись статьи представляется в редакцию по электронной почте в двух форматах pdf и tex. Статья должна быть набрана с использованием макропакета $\text{\LaTeX}$ и стиля msmb.cls, предоставляемого редакцией <http://msm.univer.omsk.su/files/msmb.zip>. Рекомендуется установить компилятор $\text{\LaTeX}$, так как именно им пользуются в редакции.

Отклонения в оформлении рукописи от приведённых правил позволяют редколлегии принять решение о снятии статьи с публикации. Статья может быть отклонена по причинам несоответствия тематике журнала или в связи с низким уровнем качества научного исследования.

В статье запрещается переопределять стандартные команды и окружения.

Нумеруемые формулы необходимо выделять в отдельную строку.

Нумерация только арабскими цифрами в порядке возрастания с единицы. Нумеровать следует только те формулы, на которые в тексте имеются ссылки.

¹Необходимы полные данные экспертов (место работы, учёная степень, должность), с указанием способа связи с ними (e-mail, телефон). Редколлегия может обратиться к одному из экспертов из предложенного списка с просьбой написать рецензию или может назначить рецензента из собственного списка.

Запрещается использовать в формулах буквы русского алфавита. Если без них никак не обойтись, то следует использовать команду `\mbox{...}`.

Все рисунки и таблицы должны иметь подпись, оформленную с помощью команды `\caption{...}`.

Файлы с рисунками необходимо представить в формате PDF или EPS (использовать редакторы векторной графики типа InkScape, Adobe Illustrator или Corel Draw).

Используйте стандартные команды переключения на готический, каллиграфический и ажурный шрифты: `\mathfrak`, `\mathcal` и `\mathbb`.

Не допускается заканчивать статью рисунком или таблицей.

В списке литературы обязательно указание следующих данных: для книг — фамилии и инициалы авторов, название книги, место издания, издательство, год издания, количество страниц; для статей — фамилии и инициалы авторов, название статьи, название журнала, год издания, том, номер (выпуск), страницы начала и конца статьи (для депонированных статей обязательно указать номер регистрации).

Кавычки в русском тексте («абвгд») должны быть угловыми, в английском — прямыми верхними кавычками ("abcdeг" или "abcdeг").

Обязательна расшифровка сокращений при первом вхождении термина. Например: ... искусственный интеллект (ИИ)...

Порядок рецензирования

Первичная экспертиза проводится главным редактором (заместителем главного редактора). При первичной экспертизе оценивается соответствие статьи тематике журнала, правилам оформления и требованиям, установленным редакцией журнала к научным публикациям.

Все статьи, поступившие в редакцию научного журнала «Математические структуры и моделирование», проходят через институт рецензирования.

Рецензент выбирается главным редактором журнала из числа членов редколлегии или ведущих специалистов по профилю данной работы.

Рецензенты уведомляются о том, что присланные им рукописи являются частной собственностью авторов и относятся к сведениям, не подлежащим разглашению. Рецензентам не разрешается делать копии статей для своих нужд.

Срок для написания рецензии устанавливается по согласованию с рецензентом.

Рецензия должна раскрывать актуальность представленного материала, степень научной новизны исследования, определять соответствие предлагаемого к публикации текста общему профилю издания и стиль изложения.

Рецензент выносит заключение о возможности опубликования статьи: «рекомендуется», «рекомендуется с учётом исправления замечаний, отмеченных рецензентом» или «не рекомендуется». В случае отрицательной рецензии редакция направляет автору мотивированный отказ, заверенный главным редактором или его заместителем.

В случае несогласия с мнением рецензента автор статьи имеет право предоставить аргументированный ответ в редакцию журнала. Статья может быть направлена на повторное рецензирование, либо на согласование в редакционную коллегию.

При наличии в рецензии рекомендаций по исправлению и доработке статьи автору направляется текст рецензии с предложением учесть их при подготовке нового варианта статьи или аргументированно (частично или полностью) их опровергнуть. Доработанная (переработанная) автором статья повторно направляется на рецензирование и рассматривается в общем порядке. В этом случае датой поступления в редакцию считается дата возвращения доработанной статьи.

После принятия редколлегией решения о допуске статьи к публикации автор информируется об этом и указываются сроки публикации.

Оригиналы рецензий хранятся в редакции в течение пяти лет.

Авторская этика

Авторы публикаций должны гарантировать, что в список авторов включены только лица, соответствующие критериям авторства (лица, внёсшие значительный вклад в работу), и что заслуживающие авторства исследователи не исключены из списка авторов.

Должны работать вместе с редакторами или издателями для скорейшего исправления своих работ в случае обнаружения в них ошибок или упущений после публикации.

Обязаны незамедлительно уведомлять редакцию в случае обнаружения ошибки в любой поданной ими на публикацию, принятой для публикации или уже опубликованной работе.

Не вправе копировать из других публикаций ссылки на работы, с которыми они сами не ознакомились; цитаты и ссылки на другие работы должны быть точными и оформленными в соответствии с предъявляемыми требованиями.

Должны ссылаться максимально правильно и точно на имеющие отношение к публикации предыдущие работы как других исследователей, так и самих авторов, обращаясь, прежде всего к первоисточнику; дословное воспроизведение собственных работ и их перефразирование неприемлемы, они могут быть использованы лишь в качестве основы для новых выводов.

Необходимо указывать авторство данных, текста, рисунков и идей, которые автор получил из других источников — они не должны представляться, как принадлежащие автору публикации; прямые цитаты из работ других исследователей должны выделяться кавычками и соответствующей ссылкой.

Должны соблюдать нормы законодательства о защите авторских прав; материалы, защищённые авторским правом (например, таблицы, цифры или крупные цитаты), могут воспроизводиться только с разрешения их владельцев.

Памятка для перевода должностей, учёных степеней и званий на английский язык

Профессор = Professor

Доцент = Associate Professor

Старший преподаватель = Assistant Professor

Преподаватель = Instructor

Ассистент = Instructor

Аспирант = Postgraduate Student или Ph.D. Student

Соискатель = Ph.D. Doctoral Candidate

Магистрант = Master's Degree Student

Студент = Student

д.ф.-м.н. = Dr.Sc. (Phys.-Math.)

к.ф.-м.н. = Ph.D. (Phys.-Math.)

д.т.н. = Dr.Sc. (Eng.)

к.т.н. = Ph.D. (Eng.)

Инженер-программист = Software Engineer

Старший/младший научный сотрудник = Senior/Junior Scientist Researcher

Электронная почта для отправки статей

lavrov@omsu.ru — зам. главного редактора (ответственный за выпуск) Д.Н. Лавров.

Научный журнал

Математические структуры и моделирование

№2(46)

Главный редактор

А.К. Гуц

Зам. глав. ред., выпускающий редактор

Д.Н. Лавров

Зам. глав. ред., технический редактор

Н.Ф. Богаченко

Корректор:

И.Н. Баловнева

Проверка корректности перевода:

А.Н. Кабанов

Адрес научной редакции

644077, Омская обл., г. Омск, пр-т Мира, д. 55а,
Омский государственный университет

E-mail: guts@omsu.ru, lavrov@omsu.ru

Электронная версия журнала:

<http://msm.univer.omsk.su>

<http://msm.omsu.ru>



Подписано в печать 20.06.2018. Формат 60 × 84 1/8.

Усл. печ. л. 19,76. Тираж 100 экз. Заказ №152.

Отпечатано на полиграфической базе издательства ОмГУ им. Ф.М. Достоевского
644077, Омская обл., г. Омск, пр-т Мира, д. 55а

ISSN 2222-8772



9 772222 877005



18046 >