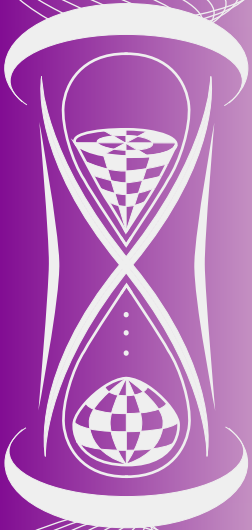


ISSN 2222-8772

**МАТЕМАТИЧЕСКИЕ
СТРУКТУРЫ
И
МОДЕЛИРОВАНИЕ**

**№ 1 (69)
2024**



**МИНИСТЕРСТВО НАУКИ
И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ**

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

**МАТЕМАТИЧЕСКИЕ
СТРУКТУРЫ
И
МОДЕЛИРОВАНИЕ**

№ 1 (69)

Омск
2024

Редакционная коллегия

- А. К. Гуц** **главный редактор**, председатель редакционной коллегии, доктор физ.-мат. наук, профессор, заслуженный работник высшей школы Российской Федерации, профессор кафедры информационно-технических дисциплин Международного инновационного университета, г. Сочи
- Д. Н. Лавров** **зам. гл. редактора**, выпускающий редактор, канд. техн. наук, доцент, проректор по образовательной деятельности и информатизации Нижневартковского государственного университета, г. Нижневартовск
- Н. Ф. Богаченко** **зам. гл. редактора**, технический редактор, канд. физ.-мат. наук, доцент, зав. кафедрой компьютерных технологий и сетей, Омский государственный университет им. Ф. М. Достоевского, г. Омск
- В. П. Голубятников** доктор физ.-мат. наук, профессор Новосибирского государственного университета, главный научный сотрудник Института математики СО РАН им. С.Л. Соболева, г. Новосибирск
- С. И. Горлов** доктор физ.-мат. наук, профессор, ректор Нижневартковского государственного университета, г. Нижневартовск
- А. Г. Гринь** доктор физ.-мат. наук, профессор, кафедра информационной безопасности, Омский государственный университет им. Ф. М. Достоевского, г. Омск
- В. А. Еровенко** доктор физ.-мат. наук, профессор, г. Минск, Республика Беларусь
- В. Zilber** Dr.Sc. (Phys.-Math.), Professor of Mathematical Logic, Mathematical Institute, University of Oxford, United Kingdom
- А. Н. Кабанов** канд. физ.-мат. наук, декан факультета цифровых технологий и кибербезопасности, Омский государственный университет им. Ф. М. Достоевского, г. Омск
- А. В. Копыльцов** доктор техн. наук, профессор, кафедра информационных систем, Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина), г. Санкт-Петербург
- А. Г. Коробейников** доктор техн. наук, профессор, зам. директора по науке Санкт-Петербургского филиала Института земного магнетизма, ионосферы и распространения радиоволн им. Н. В. Пушкова РАН, г. Санкт-Петербург
- П. А. Корчагин** доктор техн. наук, профессор, и.о. ректора Омского государственного технического университета, г. Омск
- V. Kreinovich** Ph.D. (Phys.-Math.), Professor, Computer Science Department, University of Texas at El Paso, Texas, USA
- В. А. Плетюхов** доктор физ.-мат. наук, профессор кафедры общей и теоретической физики Брестского государственного университета им. А. С. Пушкина, г. Брест, Республика Беларусь
- Л. Б. Соколинский** доктор физ.-мат. наук, профессор, зав. кафедрой системного программирования, Южно-Уральский государственный университет (национальный исследовательский университет), г. Челябинск
- А. А. Fedorenko** Ph.D. (Phys.-Math.), Researcher (CR1) at the French National Centre of Scientific Research (CNRS) Laboratoire de Physique de l'ENS-Lyon, France
- А. Jadczyk** Ph.D., Professor, Researcher, Laboratoire de Physique, Universite de Toulouse III et CNRS, France, and Ronin Institute, Montclair, New Jersey, USA

Учредитель

Федеральное государственное автономное образовательное учреждение высшего образования «Омский государственный университет им. Ф. М. Достоевского»

Журнал зарегистрирован в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций. Реестровая запись ПИ № ФС77-72200 от 15 января 2018 г.

Адрес редакции, издателя и типографии

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

Дата выхода в свет: 12.04.2024.

Тираж 25 экз.

Свободная цена.

**МАТЕМАТИЧЕСКИЕ
СТРУКТУРЫ
И
МОДЕЛИРОВАНИЕ**

Журнал основан в 1998 году. В журнале публикуются статьи, в которых излагаются результаты исследований по математике, теоретической физике, компьютерным наукам, философии и истории математики и информатики.

Все статьи журнала проходят обязательное рецензирование. Рефераты статей журнала опубликованы в «Реферативном журнале», «Mathematical Reviews» (США), «Zentralblatt für Mathematik» (Германия) и в РИНЦ (elibrary.ru). Журнал входит в Перечень рецензируемых научных изданий ВАК РФ, входящих в Международные реферативные базы данных (МРБД) и системы цитирования (Список от 31.12.2023, № 769), в которых должны быть опубликованы основные результаты диссертаций на соискание учёных степеней (Приказ Минобрнауки России от 25.07.2014 № 793). Имеет категорию К1.

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

Электронная версия журнала представлена в сети:

<http://msm.omsu.ru>

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

Подписной индекс по каталогу «Пресса России»: 94082

Электронная почта главного редактора:

aguts@mail.ru

Электронная почта выпускающего редактора:

dmtrl@ya.ru

СОДЕРЖАНИЕ

Фундаментальная математика и физика

- Д.Н. Баротов, Р.Н. Баротов. *Об одном приведении системы булевых уравнений к эквивалентной системе полиномиальных уравнений* 5
- В.В. Варламов. *Теоретико-групповое описание периодической системы элементов. IV: Групповая алгебра* 18
- А.К. Гуц. *Физическое состояние объектов прошлого и будущего* 32
- С.Л. Дерябин, А.П. Садов. *Задача Коши для двумерных течений газа, гравитирующего по Ньютонову в пространстве R^2* 42

Прикладная математика и моделирование

- А.А. Березин, С.А. Агалаков. *Применение моделей машинного обучения для диагностики фенотипов заболевания желудочно-кишечного тракта* 56
- Т.С. Катермина, А.А. Макамбеджан. *Анализ методов кластеризации для поддержки принятия решений в розничной торговле* 69
- Т.В. Леванова, А.В. Гетте. *Генетические алгоритмы для размещения наземных станций в глобальной спутниковой сети связи* 77

Компьютерные науки

- С.В. Лейхтер. *Обзор методов регистрации изображений на основе нежёстких преобразований* ... 88

Продолжение на следующей странице

Наши публикации



Информационная безопасность

- Н.Ф. Богаченко. *Инжиниринг привилегий в задаче построения ролевой политики разграничения доступа* 94
- Т.В. Вахний, С.В. Вахний. *Оптимизация выбора стратегии защиты от вредоносных атак на основе решения биматричной игры с учётом рисков* . 103
- А.И. Горев, Е.Г. Горева. *Кибератаки: некоторые подходы к системному анализу* 110
- В.Н. Семенихин. *Различные схемы и подходы к процессам журналирования* 117

Некролог

Романьков Виталий Анатольевич (1948–2023) 126

ОБ ОДНОМ ПРИВЕДЕНИИ СИСТЕМЫ БУЛЕВЫХ УРАВНЕНИЙ К ЭКВИВАЛЕНТНОЙ СИСТЕМЕ ПОЛИНОМИАЛЬНЫХ УРАВНЕНИЙ

Д.Н. Баротов¹

старший преподаватель, e-mail: DNBarotov@fa.ru

Р.Н. Баротов²

докторант, e-mail: ruzmet.tj@mail.ru

¹Финансовый университет при Правительстве Российской Федерации, Москва, Россия

²Худжандский государственный университет имени академика Б. Гафурова, Худжанд,
Таджикистан

Аннотация. В данной работе исследуется задача конструирования специального продолжения булевой функции на всё пространство \mathbb{R}^n , благодаря которому без добавления каких-либо ограничений система m булевых уравнений преобразуется в эквивалентную систему m полиномиальных уравнений. В результате исследования для любой булевой функции $f_b(x_1, x_2, \dots, x_n)$ конструируется соответствующая бесконечно дифференцируемая рациональная функция $f_s(x_1, x_2, \dots, x_n)$ такая, что

$$f_s(x_1, x_2, \dots, x_n) \in \{0, 1\} \iff \begin{cases} (x_1, x_2, \dots, x_n) \in \{0, 1\}^n \\ f_b(x_1, x_2, \dots, x_n) = f_s(x_1, x_2, \dots, x_n) \end{cases}.$$

Благодаря конструированной функции $f_s(x_1, x_2, \dots, x_n)$, во-первых, без добавления каких-либо ограничений произвольная система m булевых уравнений преобразуется в эквивалентную систему m рациональных уравнений, во-вторых, решение преобразованной эквивалентной системы рациональных уравнений сводится к задаче численной минимизации некоторой бесконечно дифференцируемой целевой функции, решаемой методами оптимизации, и к эквивалентной системе полиномиальных уравнений, решаемой и анализируемой алгоритмом F4.

Ключевые слова: продолжение булевой функции, система булевых уравнений, глобальная оптимизация, алгоритм F4, SAT.

Введение

Система булевых уравнений была важной темой исследований на протяжении почти двух столетий, и её значимость и сегодня трудно переоценить. Решение булевых уравнений проникает во многие области современной науки, такие как логическое проектирование, биология, грамматика, химия, право, медицина, спектроскопия и теория графов [1]. Многие важные задачи исследования операций можно

свести к задаче решения системы булевых уравнений. Ярким примером является задача коалиционной игры n лиц с отношением доминирования между различными стратегиями [2]. Решения булевых уравнений также служат важным инструментом при обработке псевдобулевых уравнений и неравенств и связанных с ними задач целочисленного линейного программирования [2]. В последние годы ещё одной важной и перспективной областью, в которой применяется решение системы булевых уравнений, является алгебраический криптоанализ. Для конкретного шифра алгебраический криптоанализ состоит из двух этапов: преобразовать шифр в систему полиномиальных уравнений (обычно над булевым кольцом) и решить полученную систему полиномиальных уравнений [3]. Одно из первых успешных применений решение системы булевых уравнений было продемонстрировано в работе [4]. Поэтому, с одной стороны, совершенствуются существующие методы и алгоритмы, с другой стороны, разрабатывается и адаптируется множество новых направлений и алгоритмов решения систем булевых уравнений [5–11]. Одним из таких направлений является преобразование системы булевых уравнений в систему уравнений над полем действительных чисел, поскольку в этой области известно множество методов и алгоритмов решения систем. Суть этого направления состоит в том, что система булевых уравнений преобразуется в систему уравнений над полем действительных чисел и решение ищется на множестве действительных чисел. В свою очередь преобразованная система может быть сведена к задаче численной оптимизации, что позволяет применять, анализировать и комбинировать такие методы, как алгоритм наискорейшего спуска, метод Ньютона и алгоритм координатного спуска [11–20]. Существует множество способов преобразования системы булевых уравнений в систему уравнений над полем действительных чисел [11–22]. Но одна из основных проблем, возникающих при применении этих методов, заключается в том, что преобразованная система уравнений на множестве действительных чисел может иметь множество посторонних решений, что усложняет их практическое использование [16, 18].

В настоящей работе немного уточняется результат, приведённый в [19], а именно конструируется «подходящее» продолжение (определение будет дано ниже) произвольной булевой функции на всё пространство \mathbb{R}^n , благодаря которому доказываётся, что без добавления каких-либо ограничений система m булевых уравнений и преобразованная система m рациональных уравнений на множестве действительных чисел будут эквивалентны, т. е. проблема посторонних решений полностью решается. Также доказываётся, что на основе предъявленного подходящего продолжения булевой функции $f(x_1, x_2, \dots, x_n)$ можно конструировать новую функцию, которая также является подходящим продолжением на \mathbb{R}^n функции $f(x_1, x_2, \dots, x_n)$.

1. Используемые определения и обозначения

Пусть $\mathbb{B}^n = \{(b_1, b_2, \dots, b_n) : b_1, b_2, \dots, b_n \in \{0, 1\}\}$ – множество всевозможных двоичных слов (булевых векторов) длины n .

Определение 1. Функцию вида $f_b : \mathbb{B}^n \rightarrow \mathbb{B}$ назовём булевой функцией.

Пусть $\text{supp}(f_b) = \{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n : f_b(b_1, b_2, \dots, b_n) = 1\}$ – носитель булевой

функции $f_b(x_1, x_2, \dots, x_n)$, т. е. множество всех булевых векторов, на которых булева функция $f_b(x_1, x_2, \dots, x_n)$ принимает значение 1.

Пусть $and_b(b_1, b_2, \dots, b_n) = b_1 \wedge b_2 \wedge \dots \wedge b_n$ – логическое произведение булевых переменных b_1, b_2, \dots, b_n .

Пусть $xor_b(b_1, b_2, \dots, b_n) = b_1 \oplus b_2 \oplus \dots \oplus b_n$ – логическая сумма (сумма по модулю 2) булевых переменных b_1, b_2, \dots, b_n .

Определение 2. Функцию вида $f_r : \mathbb{R}^n \rightarrow \mathbb{R}$ назовём продолжением на \mathbb{R}^n булевой функции $f_b(x_1, x_2, \dots, x_n)$, если

$$f_r(b_1, b_2, \dots, b_n) = f_b(b_1, b_2, \dots, b_n) \quad \forall (b_1, b_2, \dots, b_n) \in \mathbb{B}^n.$$

Определение 3. Функцию вида $f_s : \mathbb{R}^n \rightarrow \mathbb{R}$ назовём подходящим продолжением на \mathbb{R}^n булевой функции $f_b(x_1, x_2, \dots, x_n)$, если $f_s(x_1, x_2, \dots, x_n) \in C^\infty(\mathbb{R}^n)$ и

$$f_s(x_1, x_2, \dots, x_n) \in \{0, 1\} \iff \begin{cases} (x_1, x_2, \dots, x_n) \in \{0, 1\}^n \\ f_b(x_1, x_2, \dots, x_n) = f_s(x_1, x_2, \dots, x_n) \end{cases}.$$

2. Конструирование подходящих продолжений булевых функций

В этом разделе мы сначала построим подходящее продолжение для элементарных функций $and_b(b_1, b_2, \dots, b_n)$ и $xor_b(b_1, b_2, \dots, b_n)$, а затем на его основе построим продолжение для произвольной булевой функции. С этой целью обоснуем справедливость следующих двух лемм.

Лемма 1. Для булевой функции $and_b(b_1, b_2, \dots, b_n)$ функция вида

$$and_s(x_1, x_2, \dots, x_n) = \prod_{k=1}^n \frac{x_k^2}{2x_k^2 - 2x_k + 1} + \frac{1}{n} \cdot \sum_{k=1}^n \frac{(x_k^2 - x_k)^2}{(2x_k^2 - 2x_k + 1)^2} \quad (1)$$

является подходящим продолжением на \mathbb{R}^n .

Доказательство. Действительно, предъявленная функция $and_s(x_1, x_2, \dots, x_n)$ является подходящим продолжением на \mathbb{R}^n , для этого достаточно проверить справедливость следующих свойств:

- (a) $0 \leq and_s(x_1, x_2, \dots, x_n) \leq 1, \quad \forall (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$
- (b) Если $(x_1, x_2, \dots, x_n) \in \mathbb{B}^n$, то $and_s(x_1, x_2, \dots, x_n) \in \mathbb{B} = \{0, 1\}.$
- (c) $and_s(x_1, x_2, \dots, x_n) = 0 \iff (x_1, x_2, \dots, x_n) \in \mathbb{B}^n \setminus \{(1, 1, \dots, 1)\}.$
- (d) $and_s(x_1, x_2, \dots, x_n) = 1 \iff (x_1, x_2, \dots, x_n) = (1, 1, \dots, 1).$

(a) Первое неравенство, которое находится слева, очевидно, так как

$$\frac{x_k^2}{2x_k^2 - 2x_k + 1} \geq 0 \quad \frac{(x_k^2 - x_k)^2}{(2x_k^2 - 2x_k + 1)^2} \geq 0, \quad \forall x_k \in \mathbb{R} \text{ и } \forall k \in \{1, 2, \dots, n\}$$

и, следовательно,

$$\text{and}_s(x_1, x_2, \dots, x_n) = \prod_{k=1}^n \frac{x_k^2}{2x_k^2 - 2x_k + 1} + \frac{1}{n} \cdot \sum_{k=1}^n \frac{(x_k^2 - x_k)^2}{(2x_k^2 - 2x_k + 1)^2} \geq 0.$$

Теперь докажем второе неравенство, которое находится справа. Для этого в процессе пользуемся в том числе неравенством между средним арифметическим и средним геометрическим.

$$\begin{aligned} \text{and}_s(x_1, x_2, \dots, x_n) &= \frac{x_1^2}{2x_1^2 - 2x_1 + 1} \cdot \frac{x_2^2}{2x_2^2 - 2x_2 + 1} \cdot \dots \cdot \frac{x_n^2}{2x_n^2 - 2x_n + 1} + \\ &\quad + \frac{1}{n} \cdot \sum_{k=1}^n \frac{(x_k^2 - x_k)^2}{(2x_k^2 - 2x_k + 1)^2} = \\ &= \left(\sqrt[n]{\frac{x_1^2}{2x_1^2 - 2x_1 + 1} \cdot \frac{x_2^2}{2x_2^2 - 2x_2 + 1} \cdot \dots \cdot \frac{x_n^2}{2x_n^2 - 2x_n + 1}} \right)^n + \\ &\quad + \frac{1}{n} \cdot \sum_{k=1}^n \frac{(x_k^2 - x_k)^2}{(2x_k^2 - 2x_k + 1)^2} \leq \\ &\leq \left(\frac{1}{n} \cdot \left(\frac{x_1^2}{2x_1^2 - 2x_1 + 1} + \frac{x_2^2}{2x_2^2 - 2x_2 + 1} + \dots + \frac{x_n^2}{2x_n^2 - 2x_n + 1} \right) \right)^n + \\ &\quad + \frac{1}{n} \cdot \sum_{k=1}^n \frac{(x_k^2 - x_k)^2}{(2x_k^2 - 2x_k + 1)^2} \leq \\ &\leq \frac{1}{n} \cdot \left(\frac{x_1^2}{2x_1^2 - 2x_1 + 1} + \frac{x_2^2}{2x_2^2 - 2x_2 + 1} + \dots + \frac{x_n^2}{2x_n^2 - 2x_n + 1} \right) + \\ &\quad + \frac{1}{n} \cdot \sum_{k=1}^n \frac{(x_k^2 - x_k)^2}{(2x_k^2 - 2x_k + 1)^2} = \frac{1}{n} \cdot \sum_{k=1}^n \frac{x_k^2}{2x_k^2 - 2x_k + 1} + \frac{1}{n} \cdot \sum_{k=1}^n \frac{(x_k^2 - x_k)^2}{(2x_k^2 - 2x_k + 1)^2} = \\ &= \frac{1}{n} \cdot \sum_{k=1}^n \frac{x_k^2 \cdot (2x_k^2 - 2x_k + 1)}{(2x_k^2 - 2x_k + 1)^2} + \frac{1}{n} \cdot \sum_{k=1}^n \frac{x_k^2 \cdot (2x_k^2 - 2x_k + 1) - x_k^4}{(2x_k^2 - 2x_k + 1)^2} = \\ &= \frac{1}{n} \cdot \sum_{k=1}^n \frac{2 \cdot x_k^2 (2x_k^2 - 2x_k + 1) - x_k^4}{(2x_k^2 - 2x_k + 1)^2} = \frac{1}{n} \cdot \\ &\quad \sum_{k=1}^n \frac{2 \cdot x_k^2 (2x_k^2 - 2x_k + 1) - x_k^4 - (2x_k^2 - 2x_k + 1)^2}{(2x_k^2 - 2x_k + 1)^2} + \frac{1}{n} \cdot \\ &\quad \sum_{k=1}^n \frac{(2x_k^2 - 2x_k + 1)^2}{(2x_k^2 - 2x_k + 1)^2} = -\frac{1}{n} \cdot \sum_{k=1}^n \frac{(x_k^2 - 2x_k + 1)^2}{(2x_k^2 - 2x_k + 1)^2} + \frac{n}{n} \leq 1, \quad \forall (x_1, x_2, \dots, x_n) \in \mathbb{R}^n. \end{aligned}$$

(b) Если $(x_1, x_2, \dots, x_n) \in \mathbb{B}^n$, то $x_k \in \mathbb{B}, \forall k \in \{1, 2, \dots, n\}$. Отсюда

$$\frac{x_k^2}{2x_k^2 - 2x_k + 1} = \frac{x_k^2}{2x_k \cdot (x_k - 1) + 1} = \frac{x_k}{0 + 1} = x_k \in \mathbb{B},$$

$$\frac{(x_k^2 - x_k)^2}{(2x_k^2 - 2x_k + 1)^2} = \frac{(x_k(x_k - 1))^2}{(2x_k(x_k - 1) + 1)^2} = \frac{0^2}{(0 + 1)^2} = 0 \in \mathbb{B}, \quad \forall k \in \{1, 2, \dots, n\}.$$

Следовательно,

$$\prod_{k=1}^n \frac{x_k^2}{2x_k^2 - 2x_k + 1} + \frac{1}{n} \cdot \sum_{k=1}^n \frac{(x_k^2 - x_k)^2}{(2x_k^2 - 2x_k + 1)^2} = \text{and}_s(x_1, x_2, \dots, x_n) \in \mathbb{B}.$$

(с) Действительно,

$$\begin{aligned} \text{and}_s(x_1, x_2, \dots, x_n) &= \prod_{k=1}^n \frac{x_k^2}{2x_k^2 - 2x_k + 1} + \frac{1}{n} \cdot \sum_{k=1}^n \frac{(x_k^2 - x_k)^2}{(2x_k^2 - 2x_k + 1)^2} = 0 \iff \\ &\iff \begin{cases} \prod_{k=1}^n \frac{x_k^2}{2x_k^2 - 2x_k + 1} = 0 \\ \sum_{k=1}^n \frac{(x_k^2 - x_k)^2}{(2x_k^2 - 2x_k + 1)^2} = 0 \end{cases} \iff \begin{cases} \prod_{k=1}^n x_k^2 = 0 \\ \sum_{k=1}^n (x_k^2 - x_k)^2 = 0 \end{cases} \iff \\ &\iff \begin{cases} x_1 \cdot x_2 \cdot \dots \cdot x_n = 0 \\ x_1^2 - x_1 = 0 \\ x_2^2 - x_2 = 0 \\ \dots \dots \dots \\ x_n^2 - x_n = 0 \end{cases} \iff (x_1, x_2, \dots, x_n) \in \mathbb{B}^n \setminus \{(1, 1, \dots, 1)\}. \end{aligned}$$

(d) Действительно, в силу (a)

$$\begin{aligned} \text{and}_s(x_1, x_2, \dots, x_n) &= \prod_{k=1}^n \frac{x_k^2}{2x_k^2 - 2x_k + 1} + \frac{1}{n} \cdot \sum_{k=1}^n \frac{(x_k^2 - x_k)^2}{(2x_k^2 - 2x_k + 1)^2} = 1 \iff \\ &\iff \begin{cases} \sum_{k=1}^n \frac{(x_k^2 - 2x_k + 1)^2}{(2x_k^2 - 2x_k + 1)^2} = 0 \\ -\frac{1}{n} \cdot \sum_{k=1}^n \frac{x_k^2}{2x_k^2 - 2x_k + 1} + \prod_{k=1}^n \frac{x_k^2}{2x_k^2 - 2x_k + 1} = 0 \end{cases} \iff \\ &\iff \begin{cases} (x_k^2 - 2x_k + 1)^2 = 0, \quad \forall k \in \{1, 2, \dots, n\} \\ \frac{x_i^2}{2x_i^2 - 2x_i + 1} = \frac{x_j^2}{2x_j^2 - 2x_j + 1}, \quad \forall i, j \in \{1, 2, \dots, n\} \end{cases} \iff \\ &\iff (x_1, x_2, \dots, x_n) = (1, 1, \dots, 1). \end{aligned}$$

Лемма подробно доказана. ■

Замечание 1. Нет свойства единственности подходящего продолжения на \mathbb{R}^n булевой функции $\text{and}_b(b_1, b_2, \dots, b_n)$, поскольку, например, если $\text{and}_s(x_1, x_2, \dots, x_n)$ подходящее продолжение булевой функции $\text{and}_b(b_1, b_2, \dots, b_n)$, то $(\text{and}_s(x_1, x_2, \dots, x_n))^2$ также является подходящим продолжением на \mathbb{R}^n функции $\text{and}_b(b_1, b_2, \dots, b_n)$.

Лемма 2. Для булевой функции $xor_b(b_1, b_2, \dots, b_n)$ функция вида

$$xor_s(x_1, x_2, \dots, x_n) = \frac{1}{2} - \frac{1}{2} \cdot \prod_{k=1}^n \frac{2 - 4x_k}{1 + (1 - 2x_k)^2} \quad (2)$$

является подходящим продолжением на \mathbb{R}^n .

Доказательство. Действительно, представленная функция $xor_s(x_1, x_2, \dots, x_n)$ является подходящим продолжением на \mathbb{R}^n , для этого достаточно показать справедливость следующих свойств:

$$(a) \quad 0 \leq xor_s(x_1, x_2, \dots, x_n) \leq 1, \quad \forall (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

$$(b) \quad xor_s(x_1, x_2, \dots, x_n) \in \mathbb{B} \iff (x_1, x_2, \dots, x_n) \in \mathbb{B}^n.$$

$$(c) \quad xor_s(x_1, x_2, \dots, x_n) = 0 \iff (x_1, x_2, \dots, x_n) \in \mathbb{B}^n \text{ и } (x_1 + x_2 + \dots + x_n) - \text{чётно.}$$

$$(d) \quad xor_s(x_1, x_2, \dots, x_n) = 1 \iff (x_1, x_2, \dots, x_n) \in \mathbb{B}^n \text{ и } (x_1 + x_2 + \dots + x_n) - \text{нечётно.}$$

(a) Действительно, так как очевидно, что $(-2x_k)^2 \geq 0$ и $(2 - 2x_k)^2 \geq 0$, $\forall x_k \in \mathbb{R}$.
Из этих неравенств следует

$$-1 \leq \frac{2 - 4x_k}{1 + (1 - 2x_k)^2} \leq 1, \quad \forall x_k \in \mathbb{R}.$$

Теперь легко заметить, что из последнего неравенства следует, что

$$-1 \leq \frac{2 - 4x_1}{1 + (1 - 2x_1)^2} \cdot \frac{2 - 4x_2}{1 + (1 - 2x_2)^2} \cdot \dots \cdot \frac{2 - 4x_n}{1 + (1 - 2x_n)^2} \leq 1, \quad \forall (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

Отсюда в силу (2) получим

$$0 \leq xor_s(x_1, x_2, \dots, x_n) \leq 1, \quad \forall (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

(b) Сначала докажем в одну сторону, если $(x_1, x_2, \dots, x_n) \in \mathbb{B}^n$, то

$$\frac{2 - 4x_k}{1 + (1 - 2x_k)^2} \in \{-1, 1\}, \quad \forall k \in \{1, 2, \dots, n\}$$

и, следовательно,

$$-\frac{1}{2} \cdot \frac{2 - 4x_1}{1 + (1 - 2x_1)^2} \cdot \frac{2 - 4x_2}{1 + (1 - 2x_2)^2} \cdot \dots \cdot \frac{2 - 4x_n}{1 + (1 - 2x_n)^2} \in \left\{ -\frac{1}{2}, \frac{1}{2} \right\}.$$

Отсюда в силу (2) получим

$$xor_s(x_1, x_2, \dots, x_n) \in \mathbb{B} = \{0, 1\}, \quad \forall (x_1, x_2, \dots, x_n) \in \mathbb{B}^n.$$

Теперь докажем в обратную сторону, если $xor_s(x_1, x_2, \dots, x_n) \in \{0, 1\}$, то в силу (2)

$$\frac{2 - 4x_1}{1 + (1 - 2x_1)^2} \cdot \frac{2 - 4x_2}{1 + (1 - 2x_2)^2} \cdot \dots \cdot \frac{2 - 4x_n}{1 + (1 - 2x_n)^2} \in \{-1, 1\}.$$

Переходя к модулям, получим

$$\left| \frac{2 - 4x_1}{1 + (1 - 2x_1)^2} \cdot \frac{2 - 4x_2}{1 + (1 - 2x_2)^2} \cdot \dots \cdot \frac{2 - 4x_n}{1 + (1 - 2x_n)^2} \right| = 1.$$

Отсюда

$$\left| \frac{2 - 4x_k}{1 + (1 - 2x_k)^2} \right| = 1, \quad \forall k \in \{1, 2, \dots, n\}.$$

Заметим, что из последнего равенства следует

$$x_k \in \{0, 1\}, \quad \forall k \in \{1, 2, \dots, n\}$$

и, следовательно,

$$(x_1, x_2, \dots, x_n) \in \mathbb{B}^n.$$

(с) Сначала докажем в одну сторону, если $xor_s(x_1, x_2, \dots, x_n) = 0$, то из пункта (b) следует, что $(x_1, x_2, \dots, x_n) \in \mathbb{B}^n$. Заметим, что если $(x_1, x_2, \dots, x_n) \in \mathbb{B}^n$, то

$$\frac{2 - 4x_k}{1 + (1 - 2x_k)^2} = \frac{2 - 4x_k}{1 + (\pm 1)^2} = 1 - 2x_k = (-1)^{x_k}, \quad \forall k \in \{1, 2, \dots, n\}$$

и, следовательно,

$$\begin{aligned} 0 &= \frac{1}{2} - \frac{1}{2} \cdot \frac{2 - 4x_1}{1 + (1 - 2x_1)^2} \cdot \frac{2 - 4x_2}{1 + (1 - 2x_2)^2} \cdot \dots \cdot \frac{2 - 4x_n}{1 + (1 - 2x_n)^2} = \\ &= \frac{1}{2} - \frac{1}{2} \cdot \frac{2 - 4x_1}{1 + (\pm 1)^2} \cdot \frac{2 - 4x_2}{1 + (\pm 1)^2} \cdot \dots \cdot \frac{2 - 4x_n}{1 + (\pm 1)^2} = \\ &= \frac{1}{2} - \frac{1}{2} \cdot (1 - 2x_1) \cdot (1 - 2x_2) \cdot \dots \cdot (1 - 2x_n) = \\ &= \frac{1}{2} - \frac{1}{2} \cdot (-1)^{x_1} \cdot (-1)^{x_2} \cdot \dots \cdot (-1)^{x_n} = \frac{1}{2} - \frac{1}{2} \cdot (-1)^{x_1 + x_2 + \dots + x_n}. \end{aligned}$$

Отсюда получим, что $(x_1 + x_2 + \dots + x_n)$ – чётно.

Теперь докажем в обратную сторону, если $(x_1, x_2, \dots, x_n) \in \mathbb{B}^n$ и $(x_1 + x_2 + \dots + x_n)$ – чётно, то

$$\begin{aligned} 0 &= \frac{1}{2} - \frac{1}{2} \cdot (-1)^{x_1 + x_2 + \dots + x_n} = \frac{1}{2} - \frac{1}{2} \cdot (-1)^{x_1} \cdot (-1)^{x_2} \cdot \dots \cdot (-1)^{x_n} = \\ &= \frac{1}{2} - \frac{1}{2} \cdot (1 - 2x_1) \cdot (1 - 2x_2) \cdot \dots \cdot (1 - 2x_n) = \\ &= \frac{1}{2} - \frac{1}{2} \cdot \frac{2 - 4x_1}{1 + (\pm 1)^2} \cdot \frac{2 - 4x_2}{1 + (\pm 1)^2} \cdot \dots \cdot \frac{2 - 4x_n}{1 + (\pm 1)^2} = \\ &= \frac{1}{2} - \frac{1}{2} \cdot \frac{2 - 4x_1}{1 + (1 - 2x_1)^2} \cdot \frac{2 - 4x_2}{1 + (1 - 2x_2)^2} \cdot \dots \cdot \frac{2 - 4x_n}{1 + (1 - 2x_n)^2} = xor_s(x_1, x_2, \dots, x_n). \end{aligned}$$

(d) Справедливость этого пункта следует из пунктов (b) и (с). Лемма подробно доказана. ■

Замечание 2. Нет свойства единственности подходящего продолжения на \mathbb{R}^n булевой функции $xor_b(b_1, b_2, \dots, b_n)$. Обоснование этого аналогично обоснованию замечания 1.

3. Конструирование подходящего продолжения произвольной булевой функции

В этом разделе на основе (1) и (2) построим подходящее продолжение для произвольной булевой функции $f_b(b_1, b_2, \dots, b_n)$.

Теорема 1. Для произвольной булевой функции $f_b(b_1, b_2, \dots, b_n)$ функция вида

$$f_s(x_1, x_2, \dots, x_n) = \frac{1}{2} - \frac{1}{2} \cdot \prod_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} \frac{2 - 4 \cdot \text{and}_s(f_b(b_1, b_2, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n})}{1 + (1 - 2 \cdot \text{and}_s(f_b(b_1, b_2, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n}))^2} \quad (3)$$

является подходящим продолжением на \mathbb{R}^n , где $x_k^{b_k} = (2b_k - 1) \cdot x_k + 1 - b_k$.

Доказательство. Легко заметить, что предъявленная функция $f_s(x_1, x_2, \dots, x_n) \in C^\infty(\mathbb{R}^n)$. В силу теоремы 2 из [21] и формы СДНФ справедливо следующее вспомогательное тождество:

$$f_b(x_1, x_2, \dots, x_n) = \bigoplus_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} f_b(b_1, b_2, \dots, b_n) \wedge x_1^{b_1} \wedge x_2^{b_2} \wedge \dots \wedge x_n^{b_n}. \quad (4)$$

Остаётся доказать, что

$$\forall b \in \{0, 1\}, \quad f_s(x_1, x_2, \dots, x_n) = b \iff (x_1, x_2, \dots, x_n) \in \mathbb{B}^n \text{ и } f_b(x_1, x_2, \dots, x_n) = b.$$

Сначала докажем в одну сторону, пусть $f_s(x_1, x_2, \dots, x_n) = b \in \{0, 1\}$. Тогда в силу леммы 2 имеем $\text{and}_s(f_b(b_1, b_2, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n}) \in \{0, 1\}$, $\forall (b_1, b_2, \dots, b_n) \in \mathbb{B}^n$. Тогда в силу леммы 1 имеем $(f_b(b_1, b_2, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n}) \in \mathbb{B}^{n+1}$, $\forall (b_1, b_2, \dots, b_n) \in \mathbb{B}^n$. Отсюда следует, что $(x_1, x_2, \dots, x_n) \in \mathbb{B}^n$. В таком случае

$$\begin{aligned} & \text{and}_s(f_b(b_1, b_2, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n}) = \\ & = f_b(b_1, b_2, \dots, b_n) \cdot ((2b_1 - 1) \cdot x_1 + 1 - b_1) \cdot ((2b_2 - 1) \cdot x_2 + 1 - b_2) \cdot \\ & \dots \cdot ((2b_n - 1) \cdot x_n + 1 - b_n) = f_b(b_1, b_2, \dots, b_n) \cdot \prod_{k=1}^n (b_k \cdot x_k + (1 - b_k) \cdot (1 - x_k)) = \\ & = f_b(b_1, b_2, \dots, b_n) \wedge \bigwedge_{k=1}^n (b_k \cdot x_k + \overline{b_k} \cdot \overline{x_k}) = f_b(b_1, b_2, \dots, b_n) \wedge \bigwedge_{k=1}^n x_k^{b_k}, \quad \forall (b_1, b_2, \dots, b_n) \in \mathbb{B}^n \end{aligned}$$

и, следовательно,

$$\begin{aligned} & \frac{2 - 4 \cdot \text{and}_s(f_b(b_1, b_2, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n})}{1 + (1 - 2 \cdot \text{and}_s(f_b(b_1, b_2, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n}))^2} = \\ & = \frac{2 - 4 \cdot \text{and}_s(f_b(b_1, b_2, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n})}{1 + (\pm 1)^2} = \end{aligned}$$

$$\begin{aligned}
 &= 1 - 2 \cdot \text{and}_s (f_b(b_1, b_2, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n}) = \\
 &= 1 - 2 \cdot f_b(b_1, b_2, \dots, b_n) \wedge \bigwedge_{k=1}^n x_k^{b_k} = (-1)^{f_b(b_1, b_2, \dots, b_n) \wedge \bigwedge_{k=1}^n x_k^{b_k}}.
 \end{aligned}$$

Отсюда

$$\begin{aligned}
 f_s(x_1, x_2, \dots, x_n) &= \frac{1}{2} - \frac{1}{2} \cdot \prod_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} (-1)^{f_b(b_1, b_2, \dots, b_n) \wedge \bigwedge_{k=1}^n x_k^{b_k}} = \\
 &= \frac{1}{2} - \frac{1}{2} \cdot (-1)^{\sum_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} f_b(b_1, b_2, \dots, b_n) \wedge \bigwedge_{k=1}^n x_k^{b_k}} = \\
 &= \left(\sum_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} f_b(b_1, b_2, \dots, b_n) \wedge \bigwedge_{k=1}^n x_k^{b_k} \right) \text{ mod } 2 = \\
 &= \bigoplus_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} f_b(b_1, b_2, \dots, b_n) \wedge x_1^{b_1} \wedge x_2^{b_2} \wedge \dots \wedge x_n^{b_n}.
 \end{aligned}$$

В силу тождества (4) справедливо

$$\bigoplus_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} f_b(b_1, b_2, \dots, b_n) \wedge x_1^{b_1} \wedge x_2^{b_2} \wedge \dots \wedge x_n^{b_n} = f(x_1, x_2, \dots, x_n) = b.$$

Теперь докажем в другую сторону, пусть $(x_1, x_2, \dots, x_n) \in \mathbb{B}^n$ и $f(x_1, x_2, \dots, x_n) = b$. Тогда

$$\begin{aligned}
 b = f(x_1, x_2, \dots, x_n) &= \bigoplus_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} f_b(b_1, b_2, \dots, b_n) \wedge x_1^{b_1} \wedge x_2^{b_2} \wedge \dots \wedge x_n^{b_n} = \\
 &= \left(\sum_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} f_b(b_1, b_2, \dots, b_n) \wedge \bigwedge_{k=1}^n x_k^{b_k} \right) \text{ mod } 2 = \frac{1}{2} - \frac{1}{2} \cdot \\
 &(-1)^{\sum_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} f_b(b_1, b_2, \dots, b_n) \wedge \bigwedge_{k=1}^n x_k^{b_k}} = \frac{1}{2} - \frac{1}{2} \cdot \prod_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} (-1)^{f_b(b_1, b_2, \dots, b_n) \wedge \bigwedge_{k=1}^n x_k^{b_k}} = \\
 &= \frac{1}{2} - \frac{1}{2} \cdot \prod_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} \left(1 - 2 \cdot f_b(b_1, b_2, \dots, b_n) \wedge \bigwedge_{k=1}^n x_k^{b_k} \right) = \\
 &= \frac{1}{2} - \frac{1}{2} \cdot \prod_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} \frac{2 - 4 \cdot \text{and}_s (f_b(b_1, b_2, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n})}{1 + (\pm 1)^2} = \\
 &= \frac{1}{2} - \frac{1}{2} \cdot \prod_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} \frac{2 - 4 \cdot \text{and}_s (f_b(b_1, b_2, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n})}{1 + (1 - 2 \cdot \text{and}_s (f_b(b_1, b_2, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n}))^2}.
 \end{aligned}$$

В силу (3) справедливо

$$\begin{aligned} \frac{1}{2} - \frac{1}{2} \cdot \prod_{(b_1, b_2, \dots, b_n) \in \mathbb{B}^n} \frac{2 - 4 \cdot \text{and}_s(f_b(b_1, b_2, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n})}{1 + (1 - 2 \cdot \text{and}_s(f_b(b_1, b_2, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n}))^2} = \\ = f_s(x_1, x_2, \dots, x_n) = b. \end{aligned}$$

Теорема доказана. ■

Замечание 3. Нет свойства единственности подходящего продолжения на \mathbb{R}^n булевой функции $f_b(b_1, b_2, \dots, b_n)$. Обоснование этого аналогично обоснованию замечания 1, т. е. благодаря одному конструированному подходящему предложению можно построить бесконечное количество новых подходящих предложений.

Рассмотрим произвольную систему булевых уравнений вида

$$\begin{cases} p_1(x_1, x_2, \dots, x_n) = 0 \\ p_2(x_1, x_2, \dots, x_n) = 0 \\ p_3(x_1, x_2, \dots, x_n) = 0 \\ \dots \quad \dots \quad \dots \\ p_m(x_1, x_2, \dots, x_n) = 0 \end{cases} \quad (5)$$

Трансформируем систему (5) в соответствующую систему подходящих уравнений вида

$$\begin{cases} p_{s1}(x_1, \dots, x_n) = \frac{1}{2} - \frac{1}{2} \cdot \prod_{(b_1, \dots, b_n) \in \mathbb{B}^n} \frac{2 - 4 \cdot \text{and}_s(p_1(b_1, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n})}{1 + (1 - 2 \cdot \text{and}_s(p_1(b_1, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n}))^2} = 0 \\ p_{s2}(x_1, \dots, x_n) = \frac{1}{2} - \frac{1}{2} \cdot \prod_{(b_1, \dots, b_n) \in \mathbb{B}^n} \frac{2 - 4 \cdot \text{and}_s(p_2(b_1, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n})}{1 + (1 - 2 \cdot \text{and}_s(p_2(b_1, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n}))^2} = 0 \\ p_{s3}(x_1, \dots, x_n) = \frac{1}{2} - \frac{1}{2} \cdot \prod_{(b_1, \dots, b_n) \in \mathbb{B}^n} \frac{2 - 4 \cdot \text{and}_s(p_3(b_1, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n})}{1 + (1 - 2 \cdot \text{and}_s(p_3(b_1, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n}))^2} = 0 \\ \dots \quad \dots \quad \dots \\ p_{sm}(x_1, \dots, x_n) = \frac{1}{2} - \frac{1}{2} \cdot \prod_{(b_1, \dots, b_n) \in \mathbb{B}^n} \frac{2 - 4 \cdot \text{and}_s(p_m(b_1, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n})}{1 + (1 - 2 \cdot \text{and}_s(p_m(b_1, \dots, b_n), x_1^{b_1}, x_2^{b_2}, \dots, x_n^{b_n}))^2} = 0 \end{cases} \quad (6)$$

В свою очередь, для системы (6) построим целевую функцию вида

$$t f_s(x_1, x_2, \dots, x_n) = \sum_{k=1}^m p_{sk}(x_1, x_2, \dots, x_n). \quad (7)$$

Теперь установим связь между (5), (6) и (7) в виде утверждения и приведём соответствующее доказательство.

Утверждение 1. В \mathbb{R}^n системы (5) и (6) эквивалентны, причём (x_1, x_2, \dots, x_n) будет решением систем (5) и (6) тогда и только тогда, когда $t f_s(x_1, x_2, \dots, x_n) = 0$.

Доказательство. В одну сторону очевидно, так как функция $p_{sk}(x_1, x_2, \dots, x_n)$ является одним из продолжений булевой функции $p_k(x_1, x_2, \dots, x_n)$, $\forall k \in \{1, 2, \dots, m\}$ и, следовательно, множество решений системы (5) является подмножеством множества решений системы (6).

В другую сторону, пусть $(r_1, r_2, \dots, r_n) \in \mathbb{R}^n$ произвольное решение системы (6). Тогда в силу лемм 1, 2 и теоремы 1 $(r_1, r_2, \dots, r_n) \in \mathbb{B}^n$ и, следовательно, (r_1, r_2, \dots, r_n) является решением системы (5), так как $p_{sk}(r_1, r_2, \dots, r_n) = p_k(r_1, r_2, \dots, r_n)$, $\forall k \in \{1, 2, \dots, m\}$. Следовательно, множество решений системы (6) в \mathbb{R}^n является подмножеством множества решений системы (5).

В силу лемм 1, 2 и теоремы 1 справедливо неравенство

$$0 \leq p_{sk}(x_1, x_2, \dots, x_n) \leq 1, \quad \forall k \in \{1, 2, \dots, m\} \quad \text{и} \quad \forall (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$$

и, следовательно, (x_1, x_2, \dots, x_n) будет решением систем (5) и (6) $\iff t f_s(x_1, x_2, \dots, x_n) = 0$. Утверждение доказано. ■

Замечание 4. Система рациональных уравнений (6) путём умножения на общий знаменатель элементарно может быть сведена к эквивалентной системе полиномиальных уравнений вида

$$\begin{cases} h_1(x_1, x_2, \dots, x_n) = 0 \\ h_2(x_1, x_2, \dots, x_n) = 0 \\ h_3(x_1, x_2, \dots, x_n) = 0 \\ \dots \quad \dots \quad \dots \\ h_m(x_1, x_2, \dots, x_n) = 0 \end{cases}, \quad (8)$$

где многочлен $h_k(x_1, x_2, \dots, x_n)$ равен числителю несократимой дроби, равной $p_{sk}(x_1, x_2, \dots, x_n)$, $\forall k \in \{1, 2, \dots, m\}$. В свою очередь, система (8) может быть проанализирована и решена алгоритмом F4 [8, 9].

Заключение

В результате исследования построено $f_s(x_1, x_2, \dots, x_n)$ – подходящее, т. е. специальное бесконечно дифференцируемое рациональное продолжение на всё пространство \mathbb{R}^n произвольной булевой функции $f_b(x_1, x_2, \dots, x_n)$, имеющее следующее важное свойство

$$f_s(x_1, x_2, \dots, x_n) \in \{0, 1\} \iff \begin{cases} (x_1, x_2, \dots, x_n) \in \{0, 1\}^n \\ f_b(x_1, x_2, \dots, x_n) = f_s(x_1, x_2, \dots, x_n) \end{cases}.$$

Благодаря этому аргументировано, что система m булевых уравнений без добавления каких-либо ограничений может быть трансформирована в эквивалентную систему m рациональных уравнений, т. е. для таких задач проблема посторонних решений, возникающих при трансформации систем, полностью решена. Также доказано, что на основе предъявленного подходящего продолжения булевой функции $f_b(x_1, x_2, \dots, x_n)$ можно конструировать новую функцию, которая также является подходящим продолжением на \mathbb{R}^n функции $f_b(x_1, x_2, \dots, x_n)$.

Полученный результат также может быть применён при решении систем уравнений, заданных одновременно как математическими, так и логическими операциями.

Литература

1. Brown F.M. Boolean Reasoning: The logic of Boolean Equations. Boston: Kluwer Academic Publishers, 1990.
2. Hammer P.L., Rudeanu S. Boolean Methods in Operations Research and Related Areas. Berlin: Springer Verlag, 1968.
3. Bard G.V. Algorithms for Solving Linear and Polynomial Systems of Equations over Finite Fields, with Applications to Cryptanalysis. College Park, MD, USA: University of Maryland, 2007.
4. Faugere J.C., Joux A. Algebraic cryptanalysis of hidden field equation (HFE) cryptosystems using Gröbner bases // Annual International Cryptology Conference. Berlin; Heidelberg: Springer, 2003. P. 44–60.
5. Armknecht F. Improving Fast Algebraic Attacks // International Workshop on Fast Software Encryption. Berlin; Heidelberg, Germany: Springer, 2004. P. 65–82.
6. Bardet M., Faugère J.-C., Salvy B., Spaenlehauer P.-J. On the complexity of solving quadratic boolean systems // J. Complex. 2013. Vol. 29. P. 53–75. DOI: 10.1016/j.jco.2012.07.001.
7. Courtois N. Fast Algebraic Attacks on Stream Ciphers with Linear Feedback // CRYPTO 2003 / Ed. D. Boneh. Berlin; Heidelberg, Germany: Springer, 2003. P. 176–174. (Lecture Notes in Computer Science, vol. 2729.)
8. Faugere J.C. A new efficient algorithm for computing Gröbner bases (F4) // J. Pure Appl. Algebra. 1999. Vol. 139. P. 61–88. DOI: 10.1016/S0022-4049(99)00005-5.
9. Faugere J.C. A new efficient algorithm for computing Gröbner bases without reduction to zero (F5) // Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation, Lille, France, 7–10 July 2002. P. 75–83.
10. Liu M., Lin, D., Pei D. Fast algebraic attacks and decomposition of symmetric Boolean functions // IEEE Trans. Inf. Theory. 2011. Vol. 57. P. 4817–4821. DOI: 10.1109/TIT.2011.2145690.
11. Abdel-Gawad A.H., Atiya A.F., Darwish N.M. Solution of systems of Boolean equations via the integer domain // Inform. Sci. 2010. Vol. 180. P. 288–300. DOI: 10.1016/j.ins.2009.09.010.
12. Gu J. How to Solve Very Large-Scale Satisfiability (VLSS) Problems: Technical Report UCECETR-90-002. Calgary, AB, Canada: University of Calgary, 1990.
13. Gu J. On optimizing a search problem // Artificial Intelligence Methods and Applications / Ed. N.G. Bourbakis. Singapore: World Scientific Publishers, 1992.
14. Gu J. Global optimization for satisfiability (SAT) problem // IEEE Trans. Knowl. Data Eng. 1994. Vol. 6. P. 361–381. DOI: 10.1109/69.334864.
15. Gu J., Gu Q., Du D. On optimizing the satisfiability (SAT) problem // J. Comput. Sci. Technol. 1999. Vol. 14. P. 1–17. DOI: 10.1007/BF02952482.
16. Баротов Д.Н., Музафаров Д.З., Баротов Р.Н. Об одном методе решения систем булевых алгебраических уравнений // Современная математика и концепции инновационного математического образования. 2021. Т. 8, № 1. С. 17–23.
17. Barotov D., Osipov A., Korchagin S., Pleshakova E., Muzafarov D., Barotov R., Serdechnyy D. Transformation Method for Solving System of Boolean Algebraic Equations // Mathematics. 2021. Vol. 9. Art. 3299. DOI: 10.3390/math9243299.
18. Barotov D.N., Barotov, R.N. Polylinear Transformation Method for Solving Systems of Logical Equations // Mathematics. 2022. Vol. 10. Art. 918. DOI: 10.3390/math10060918.

19. Barotov D.N., Barotov R.N., Soloviev V., Feklin V., Muzafarov D., Ergashboev T., Egamov K. The Development of Suitable Inequalities and Their Application to Systems of Logical Equations // Mathematics. 2022. Vol. 10. Art. 1851. DOI: 10.3390/math10111851.
20. Barotov D.N. Target Function without Local Minimum for Systems of Logical Equations with a Unique Solution // Mathematics. 2022. Vol. 10. Art. 2097. DOI: 10.3390/math10122097.
21. Баротов Д.Н., Баротов Р.Н. Полилинейные продолжения некоторых дискретных функций и алгоритм их нахождения // Вычислительные методы и программирование. 2023. № 24. С. 10–23. DOI: 10.26089/NumMet.v24r102.
22. Баротов Д.Н. Выпуклое продолжение булевой функции и его приложения // Дискретный анализ и исследование операций. 2024. Т. 31, № 1. С. 5–21. (Принята в печать)

ON THE REDUCTION OF A SYSTEM OF BOOLEAN EQUATIONS TO AN EQUIVALENT SYSTEM OF POLYNOMIAL EQUATIONS

D.N. Barotov¹

Assistant Professor, e-mail: DNBarotov@fa.ru

R.N. Barotov²

Doctoral Student, e-mail: ruzmet.tj@mail.ru

¹Financial University under the Government of the Russian Federation, Moscow, Russia

²Khujand State University named after Academician Bobojon Gafurov, Khujand, Tajikistan

Abstract. In this paper, we study the problem of constructing a special continuation of a Boolean function to the entire space \mathbb{R}^n , thanks to which, without adding any restrictions, a system of m Boolean equations is transformed into an equivalent system of m polynomial equations. As a result of the study, for any Boolean function $f_b(x_1, x_2, \dots, x_n)$, a corresponding infinitely differentiable rational function $f_s(x_1, x_2, \dots, x_n)$ is constructed such that

$$f_s(x_1, x_2, \dots, x_n) \in \{0, 1\} \iff \begin{cases} (x_1, x_2, \dots, x_n) \in \{0, 1\}^n \\ f_b(x_1, x_2, \dots, x_n) = f_s(x_1, x_2, \dots, x_n) \end{cases} .$$

Thanks to the constructed function $f_s(x_1, x_2, \dots, x_n)$, firstly, without adding any restrictions, an arbitrary system of m Boolean equations is transformed into an equivalent system of m rational equations, and secondly, the solution of the transformed an equivalent system of rational equations is reduced to the problem of numerical minimization of some infinitely differentiable target function, solved by optimization methods, and to an equivalent system of polynomial equations, solved and analyzed by the F4 algorithm.

Keywords: continuation of a Boolean function, system of Boolean equations, global optimization, F4 algorithm, SAT.

Дата поступления в редакцию: 16.11.2023

ТЕОРЕТИКО-ГРУППОВОЕ ОПИСАНИЕ ПЕРИОДИЧЕСКОЙ СИСТЕМЫ ЭЛЕМЕНТОВ. IV: ГРУППОВАЯ АЛГЕБРА

В.В. Варламов

д.ф.-м.н., e-mail: varlamov@subsiu.ru

Сибирский государственный индустриальный университет, Новокузнецк, Россия

Аннотация. Рассматривается структура групповой алгебры конформной группы (группы, лежащей в основании теоретико-группового описания периодической системы химических элементов) в рамках двукратного накрытия. Изучается водородная реализация подалгебры Картана и генераторов Вейля групповой алгебры.

Ключевые слова: периодический закон, таблица Менделеева, конформная группа, групповая алгебра, подалгебра Картана, генераторы Вейля.

1. Введение

Книга природы написана на языке математики.

Галилео Галилей

Общеизвестно, что наука только тогда достигает совершенства, когда начинает пользоваться математикой (К. Маркс). В полной мере это относится к химии, где периодический закон является наиболее важным обобщением этой науки, а периодическая таблица Менделеева называется Е. Scerri «иконой современной химии» [1]. Сразу же после открытия Д.И. Менделеевым периодической системы химических элементов стали предприниматься попытки математического описания (математизации) периодического закона (об истории вопроса см. [2]). Неудивительно, что наиболее подходящей математической структурой для описания явления периодичности, т. е. повторяемости (цикличности), оказалась теория групп.

Теоретико-групповые методы изучения периодической системы были предложены независимо несколькими авторами в начале 1970-х гг. В 1971 г. появляется работа Ю.Б. Румера и А.И. Фета [3], в которой было отмечено поразительное сходство между строением системы химических элементов и строением энергетического спектра атома водорода¹. Это сходство объясняется в [3] в рамках представления

¹Э. 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. Видимо, именно из-за отсутствия теоретического обоснования на тот момент (1936 г.) он опубликовал её в виде справочного материала в [4, с. 585].

В.А. Фока F [7] для спинорной группы $\mathbf{Spin}(4)$ (двухлистная накрывающая группы $\mathbf{SO}(4)$). Однако главным недостатком описания в [3] является приводимость представления F , что не позволяло рассматривать систему как «элементарную» в смысле групповой механики. В 1972 г. Б.Г. Конопельченко [8] расширил представление Фока F до представления F^+ конформной группы, устранив тем самым указанный выше недостаток. В том же 1972 г. появляется статья А.О. Барута [9] о групповой структуре периодической системы в рамках конформной группы $\mathbf{SO}(4, 2)$. Барут вводит квантовые числа $\nu, \lambda, \mu_\lambda, \mu_\sigma$ группы $\mathbf{SO}(4, 2)$ с целью рассмотрения химических элементов как различных *состояний* единой *квантовой системы*, которая, в свою очередь, рассматривается как своего рода сверхчастица. Следуя предложению, данному позднее С.Е. Вульфманом [11] в 1978 г., эта псевдочастица, спектр которой является атомным супермультиплетом Барута, будет обозначаться названием *барутон*. Различные состояния (или элементы) Барут представляет кет-векторами $|\alpha\rangle, |\beta\rangle, |\gamma\rangle, \dots$, которые образуют базис бесконечномерного гильбертова пространства. В рамках теоретико-группового описания различные химические элементы рассматриваются как *бесструктурные* частицы, при этом предполагается, что атомы являются *несоставными*, а поэтому их внутреннюю динамику можно игнорировать³.

Одновременно с этими публикациями появляются работы О. Новаро с соавторами [16, 17], где группой симметрии периодической системы предлагается группа $G_{NB} = \mathbf{SU}(2) \otimes \mathbf{SU}(2) \otimes \mathbf{SU}(2)$, образованная тремя взаимно коммутирующими угловыми моментами $\mathbf{J}_1, \mathbf{J}_2, \mathbf{J}_3$. Группа Новаро–Беррондо G_{NB} допускает следующую редукцию:

$$\mathbf{SU}(2) \otimes \mathbf{SU}(2) \otimes \mathbf{SU}(2) \supset \mathbf{O}(4) \supset \mathbf{SO}(3).$$

Неприводимые представления группы G_{NB} имеют вид (j_1, j_2, j_3) , где j_1, j_2, j_3 яв-

При рассмотрении истории возникновения правила Маделунга $(n + l, n)$ возникает множество приоритетных вопросов. Как отмечает В.Н. Островский [5], трудно проследить происхождение этого правила, которое выглядит как своего рода научный фольклор. Так, в 1930 г. В. Карапетов использовал это правило для предсказания трансурановых элементов до $Z = 124$ включительно [6].

²Чтобы не быть предвзятым по отношению к квантовым числам n, l, m_l и m_s механической (планетарной) модели Бора, описывающей водородоподобные системы, Барут намеренно ввёл символы $\nu, \lambda, \mu_\lambda$ и μ_σ , имеющие, прежде всего, теоретико-групповой (не механический) смысл, хотя диапазон их изменения такой же, как у «водородных» квантовых чисел. Островский аналогичным образом проводит различие между обычными («водородными») квантовыми числами и абстрактными $\mathbf{SO}(4, 2)$ -символами, обозначая последние знаком тильды: $\tilde{n}, \tilde{l}, \tilde{m}_l, \tilde{m}_s$ [10]. Этот символизм призван подчеркнуть холистическую трактовку теоретико-группового подхода в отличие от механистического редукционизма модели Бора, в которой квантовые числа соответствуют радиальным и орбитальным движениям «составных частей» атома за исключением квантового числа m_s , не имеющего, как известно, классического аналога, что лишний раз указывает на паллиативный характер модели Бора.

³Под «бесструктурностью» здесь не следует понимать отсутствие какой-либо структуры вообще. Прежде всего, под этим понимается отрицание структуры редукционистского плана в виде механических моделей (планетарная модель Резерфорда–Бора, модель кварков), привнесённых, как говорил Гейзенберг, из «репертуара классической физики», т. е. моделей, адекватных на уровне макрофизики, но теряющих свой смысл и затемняющих существо дела при переносе их на микроуровень. В рамках теоретико-группового подхода реализуется структура холистического плана. А именно, различные состояния, являющиеся *циклическими векторами* \mathbb{K} -гильбертова пространства, имеют структуру тензорного произведения [12–15]. Эта структура задаёт динамическую связь между спином, зарядом и массой.

ляются собственными значениями операторов Казимира J_1^2 , J_2^2 , J_3^2 и принимают целые и полуцелые значения. Физически допустимые представления группы G_{NB} имеют два вида: $(j, j, 0)$ и $(j, 0, j)$, $j = 0, 1, \frac{1}{2}, \dots$. Это следствие представления Фока $F(j_1 = j_2)$ для группы $O(4)$, входящей в редукционную цепочку для G_{NB} .

Уже из первых работ в этом направлении появляется важное различие между двумя теоретико-групповыми подходами. Первым на это различие указал В.Н. Островский [10]. Исторически сложилось так, что единственной изучаемой (методами теории групп) квантовой системой был атом водорода, гамильтониан которого был точно известен. Когда теория групп начала применяться в атомной физике, это было типичным случаем. Следуя терминологии Островского [10], будем называть это *подходом атомной физики (ПАФ)*. Однако, когда дело доходит до периодической системы, гораздо сложнее построить гамильтониан, не говоря уже об изучении его симметрии. *Подход элементарных частиц (ПЭЧ)*, основы которого заложены в работах Румера, Фета и Барута, опирается на аналогию с группами динамических («внутренних») симметрий физики элементарных частиц, таких как $SU(2)$ (изотопический спин), $SU(3)$ и $SU(6)$. В этом подходе химические элементы рассматриваются как различные состояния некоторой субстанции: «атомной материи» Барута⁴ [9] или «бесструктурной частицы с внутренними степенями свободы» Румера–Фета [3].

Все дальнейшие теоретико-групповые обобщения были связаны с попытками теоретического объяснения так называемых атомных *магических чисел*, описывающих удвоение периодов: 2, 8, 8, 18, 18, 32, 32, ... Как отмечает Р.-О. Löwdin [18], отсутствие теоретического объяснения удвоения периодов (имеющее место до сих пор, см. [1, 19]), равнозначно отсутствию теоретического понимания периодической системы химических элементов в целом⁵. Удвоение периодов означает, что всё многообразие химических элементов естественным образом распадается на два множества с суммой $(n+l)$, чётной или нечётной, где n и l – главное и орбитальное квантовые числа соответственно. В результате элементы из одного и того же (чётного или нечётного) множества химически более схожи, чем элементы из разных множеств

⁴Поскольку в рамках ПЭЧ различные состояния образуют *единую квантовую систему*, то, как следствие, возможны *квантовые переходы* между состояниями (трансмутация элементов). В этом контексте «атомную материю» Барута следует понимать как «первичную материю» (*prima materia*).

⁵Согласно широко распространённому заблуждению, планетарная модель Бора объясняет периодическую систему Менделеева. Однако Бор выводил электронные конфигурации не из квантовой теории, а исходя из известных химических и спектроскопических свойств элементов. Более того, интерпретация структуры периодической системы на основе последовательности заполнения электронных атомных орбиталей в соответствии с их относительными энергиями ε_{nl} весьма и весьма приближённа. Универсальной последовательности орбитальных энергий ε_{nl} не существует, к тому же такая последовательность не определяет полностью порядок заселения атомных орбиталей электронами, поскольку необходим учёт конфигурационных взаимодействий (наложение конфигураций в многоконфигурационном приближении). Неизвестна причина повторения сходных электронных конфигураций атомов (более подробно см. [20]). В результате модель Бора может воспроизвести (аппроксимировать) первоначальное открытие Менделеева только с помощью математических приближений (в рамках одноэлектронного приближения Хартри) – она не может предсказать (объяснить) периодическую систему. Как следствие, широко распространённое представление о сводимости (редукции) химии к физике ставится под сомнение [21, 22]. В связи с этим в последнее время в журнале *Foundations of Chemistry* возникла довольно широкая дискуссия об онтологическом статусе химии.

[23–25]. Барут пытался объяснить удвоение периодов при помощи редукции представлений h конформной группы $SO(4, 2)$ относительно подгруппы $SO(3, 2)$ (группа анти-де Ситтера), основанной на следующей цепочке:

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

согласно которой представление h распадается на сумму $h = h_e \oplus h_o$, где h_e соответствует чётному $(n + l)$, h_o – нечётному $(n + l)$. Островский, критикуя схему Барута, отмечает, что согласно этой редукции подгруппа $O(4)$ полностью теряет своё значение. Отсюда следует, что квантовое число n , которое непосредственно связано с $O(4)$, не появляется в этой схеме. Однако это квантовое число является существенным для описания периодической системы⁶. Более того, с группой $O(4)$ связано первое применение теории групп в квантовой механике. В статье 1926 г. [27] Паули использовал матричную механику Гейзенберга для получения спектра атома водорода. Помимо углового момента \mathbf{L} , Паули также ввёл квантовомеханический аналог \mathbf{A} классического вектора Лапласа–Рунге–Ленца. Инвариантность гамильтониана относительно этих операторов (\mathbf{L} и \mathbf{A}) оказалась достаточной для объяснения полного вырождения спектра водорода. Кроме того, соответствующая алгебра может быть идентифицирована как алгебра Ли группы вращений в четырёх измерениях, изоморфная специальной ортогональной группе $SO(4)$, что впоследствии было строго установлено В. Фоком [28] (и далее В. Баргманом [29]).

В работах А.И. Фета [30, 31] удвоение периодов интерпретируется посредством включения циклической группы \mathbb{Z}_2 (т. е. группы перестановок двух элементов) в динамическую группу:

$$G_F = O(4, 2) \otimes SU(2) \otimes \mathbb{Z}_2.$$

Далее, в 1981 г. Островский [32] вводит группу

$$G_O = O(4, 2) \otimes SU(2)_S \otimes SU(2)_T.$$

Её подгруппа $O(4) \otimes SU(2)_S \otimes SU(2)_T$ содержит симметрию $O(4)$, которая приводит к представлениям размерности n^2 . Посредством расширения группы $O(4)$ до $O(4) \otimes SU(2)_S$ размерности представлений удваиваются до $2n^2$. Нижний индекс S здесь указывает на физическое происхождение группы $SU(2)$ от электронного спина $m_s = \pm 1/2$. Островский называл это «горизонтальное» удвоение длин периодов *спиновым удвоением*. «Вертикальное» удвоение длин периодов, известное как фактическое *удвоение периодов* в периодической системе, было сформулировано Островским в теоретико-групповой форме путём введения второй группы $SU(2)$, обозначаемой $SU(2)_T$ и формально аналогичной группе изоспина. Это приводит к двум копиям представлений группы $O(4, 2) \otimes SU(2)_S$, которые реализуются в двух различных гильбертовых пространствах \mathcal{H}_+ и \mathcal{H}_- . Островский вводит три оператора T_+ , T_- и T_3 алгебры $\mathfrak{su}(2)_T$, где оператор T_3 действует как генератор Картана,

⁶Удвоение периодов, предложенное Барутом посредством двух различных представлений группы анти-де Ситтера $SO(3, 2)$, О. Новаро [26] называет «фатальным недостатком», поскольку получающиеся размерности не соответствуют магическим числам. Сам Новаро пытался объяснить удвоение периодов посредством различия двух типов представлений $(j, j, 0)$ и $(j, 0, j)$ группы G_{NB} [26].

различающий состояния из обоих подпространств \mathcal{H}_+ и \mathcal{H}_- , а лестничные операторы T_{\pm} действуют как операторы сдвига между \mathcal{H}_+ и \mathcal{H}_- . Сравнение с группой Фета G_F , где удвоение периодов задаётся циклической группой \mathbb{Z}_2 , показывает, что конструкция G_F не приводит к использованию лестничных операторов для соединения двух непересекающихся представлений группы $O(4, 2) \otimes SU(2)_S$, как это имеет место в конструкции Островского G_O . В своих последующих работах [33, 34] Фет использует по сути такую же группу $SO(4, 2) \otimes SU(2) \otimes SU(2)'$, как у Островского.

В настоящей статье, являющейся продолжением серии работ [35–38], изучается структура групповой алгебры $\mathfrak{so}(4, 2)$ конформной группы $SO(4, 2)$, группы, как показало время, наиболее подходящей для изучения периодической системы химических элементов. За исключением группы Новаро–Беррондо G_{NB} , эта группа является главной составной частью группы Фета G_F и группы Островского G_O . По этой причине групповая алгебра $\mathfrak{so}(4, 2)$ описывает основные структурные свойства периодической системы. Прежде всего, алгебра Ли $\mathfrak{so}(4, 2)$ является алгеброй третьего ранга, что делает возможным графическую визуализацию корневой и весовой диаграмм⁷. Графический аспект чрезвычайно важен, поскольку первой формой, в которой является периодический закон, есть всем известное его табличное представление. В книге Е.Г. Мазурса [40] (см. также книгу Д.В. ван Спронсона [41]) приведено более 700 различных графических представлений периодической системы Менделеева. Сам Менделеев, как известно, предпочитал трёхмерную спиралевидную конструкцию (наподобие конструкции Шанкортау). Структура алгебры $\mathfrak{so}(4, 2)$ определяется выделением подалгебры Картана и последующим построением генераторов Вейля, что позволяет задать базис Картана–Вейля алгебры $\mathfrak{so}(4, 2)$ и построить соответствующую корневую и весовую диаграммы. В п. 2 в краткой форме рассматриваются все необходимые для дальнейшего исследования групповой алгебры $\mathfrak{so}(4, 2)$ теоретические конструкции: подалгебра Картана, генераторы Вейля и инварианты Казимира. В п. 3 используется представление Барута для генераторов алгебры $\mathfrak{so}(4, 2)$ как наиболее удобное для последующей редукции $\mathfrak{so}(4, 2)$ к её подалгебрам $\mathfrak{so}(3, 1) \simeq \mathfrak{sl}(2, \mathbb{C})$, $\mathfrak{so}(4)$ и $\mathfrak{so}(2, 2)$. При этом используется переход к двукратному накрытию конформной группы $SO(4, 2)$, которое изоморфно группе $SU(2, 2)$ ⁸. Унитарные представления группы $SU(2, 2)$, следуя методу Томаса [43], изучались в работах [44–49] главным образом относительно максимальной компактной подгруппы $K = SU(2) \otimes SU(2) \otimes U(1)$. Конечномерные представления группы $SU(2, 2)$ определяются в базисе Яо [47] относительно подгруппы K .

⁷Для алгебр Ли ранга $r > 3$ графическая реализация корневых и весовых диаграмм становится практически невозможной. Альтернативным методом построения таких диаграмм для алгебр Ли любого ранга является метод диаграмм Дынкина [39].

⁸С чисто алгебраической точки зрения более адекватно рассмотрение конформной группы с обратной сигнатурой, т. е. $SO(2, 4)$, поскольку в этом случае двулистное накрытие (спинорная группа) $\mathbf{Spin}_+(2, 4) \simeq SU(2, 2)$ имеет кватернионное кольцо деления $\mathbb{K} \simeq \mathbb{H}$ в отличие от $\mathbf{Spin}_+(4, 2)$ с вещественным кольцом $\mathbb{K} \simeq \mathbb{R}$, см. [42].

2. Подалгебра Картана и генераторы Вейля

Пусть \mathfrak{g} – алгебра Ли размерности n , образованная генераторами \mathbf{X}_i ($i = 1 \leftarrow n$), которые удовлетворяют перестановочным соотношениям

$$[\mathbf{X}_i, \mathbf{X}_j] = \sum_{k=1}^n f_{ijk} \mathbf{X}_k \equiv f_{ijk} \mathbf{X}_k,$$

где f_{ijk} – структурные константы. m генераторов \mathbf{X}_i образуют *базис* \mathcal{B} алгебры Ли \mathfrak{g} .

Определение 1. (Максимальная абелева подалгебра.) Абелева подалгебра \mathfrak{K} алгебры Ли \mathfrak{g} называется *максимальной*, когда нет дополнительных элементов алгебры \mathfrak{g} , которые коммутируют со всеми элементами подалгебры \mathfrak{K} .

Подалгебра \mathfrak{K} более известна как *подалгебра Картана* алгебры \mathfrak{g} , а число m элементов подалгебры Картана называется *рангом* алгебры Ли \mathfrak{g} .

Определение 2. (Подалгебра Картана алгебры Ли.) Пусть \mathfrak{g} – n -мерная алгебра Ли, тогда множество всех взаимно коммутирующих базисных элементов $\{\mathbf{X}_i = \mathbf{H}_i\}$ ($i = 1 \rightarrow m$) алгебры \mathfrak{g} образует базис максимальной абелевой подалгебры $\mathfrak{K} \subset \mathfrak{g}$.

Определение 3. (Ранг алгебры Ли.) Размерность m подалгебры Картана $\mathfrak{K} \subset \mathfrak{g}$ определяет *ранг* алгебры Ли \mathfrak{g} .

Элементы \mathbf{H}_i подалгебры Картана \mathfrak{K} называются *генераторами Картана* или *элементами Картана*. Генераторы Картана удовлетворяют перестановочным соотношениям

$$[\mathbf{H}_i, \mathbf{H}_j] = 0, \quad \forall i, j = 1, \dots, m. \quad (1)$$

Это означает, что все \mathbf{H}_i одновременно диагонализированы. Обозначая их собственные значения через h_i , получим

$$\mathbf{H}_i |h_1, h_2, \dots, h_m\rangle = h_i |h_1, h_2, \dots, h_m\rangle, \quad \forall i = 1 \rightarrow m.$$

Собственные значения h_i называются *весами*. h_i можно рассматривать как компоненты m -мерного вектора \mathbf{h} , который называется *весовым вектором*. Веса подалгебры Картана используются как квантовые числа для обозначения данного мультиплета.

Далее, из оставшихся генераторов \mathbf{X}_i алгебры \mathfrak{g} ($i = 1 \rightarrow n - m$), не являющихся элементами подалгебры Картана \mathfrak{K} , образуем линейные комбинации, которые, в свою очередь, образуют множество *повышающих* и *понижающих* операторов (*лестничные операторы* \mathbf{E}_α – *генераторы Вейля* или *элементы Вейля*). Наряду с генераторами Картана \mathbf{H}_i , они составляют *базис Картана–Вейля* $\{\mathbf{H}_i, \mathbf{E}_\alpha\}$ алгебры \mathfrak{g} .

Таким образом, алгебра Ли \mathfrak{g} может быть разложена в прямую сумму, состоящую из подалгебры Картана \mathfrak{K} (m генераторов \mathbf{H}_i) и $(n - m)$ одномерных подалгебр \mathfrak{W}_α , образуемых генераторами Вейля \mathbf{E}_α :

$$\mathfrak{g} = \mathfrak{K} \bigoplus_{\alpha=1}^{n-m} \mathfrak{W}_\alpha = \mathfrak{K} \oplus \mathfrak{W}_1 \oplus \mathfrak{W}_2 \oplus \dots \oplus \mathfrak{W}_{n-m}.$$

Генераторы \mathbf{H}_i и \mathbf{E}_α удовлетворяют перестановочным соотношениям

$$[\mathbf{H}_i, \mathbf{E}_\alpha] = \alpha_i \mathbf{E}_\alpha, \quad \forall i = 1 \rightarrow m, \alpha = 1 \rightarrow n - m. \quad (2)$$

Соотношения (2) могут быть записаны следующим образом:

$$\begin{bmatrix} [\mathbf{H}_1, \mathbf{E}_\alpha] \\ [\mathbf{H}_2, \mathbf{E}_\alpha] \\ \vdots \\ [\mathbf{H}_m, \mathbf{E}_\alpha] \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix} \mathbf{E}_\alpha = \alpha \mathbf{E}_\alpha.$$

Различные α_i называются *корнями* генератора \mathbf{E}_α . Множество корней α_i может рассматриваться как совокупность компонент вектора α , называемого *корневым вектором*, который принадлежит m -мерному *корневому пространству*. Корневой вектор для каждого генератора Вейля \mathbf{E}_α может быть изображён на диаграмме, называемой *диаграммой Вейля*, размерность которой равна рангу m алгебры Ли \mathfrak{g} . Следует отметить, что в силу (1) все генераторы Картана \mathbf{H}_i имеют корни $\alpha_i = 0$ и, следовательно, располагаются в центре диаграммы Вейля.

В общем случае стандартная форма коммутационных соотношений для генераторов алгебры Ли \mathfrak{g} записывается следующим образом:

$$\begin{aligned} [\mathbf{H}_i, \mathbf{H}_j] &= 0, & \forall i, j = 1, \dots, m; \\ [\mathbf{H}_i, \mathbf{E}_\alpha] &= \alpha_i \mathbf{E}_\alpha, & \forall i = 1 \rightarrow m, \alpha = 1 \rightarrow n - m; \\ [\mathbf{E}_\alpha, \mathbf{E}_{-\alpha}] &= \alpha^i \mathbf{H}_i; \\ [\mathbf{E}_\alpha, \mathbf{E}_\beta] &= N_{\alpha\beta}^\gamma \mathbf{E}_\gamma, & \beta \neq -\alpha. \end{aligned} \quad (3)$$

Величины $N_{\alpha\beta}^\gamma$ также могут быть выражены через корневые векторы, так что мы знаем алгебру \mathfrak{g} , если известны её корни. Эти корни обладают свойством

$$\sum_{\alpha} \alpha_i \alpha_j = \delta_{ij},$$

где α может принимать только $(n - m)$ значений:

$$\alpha = \pm 1, \pm 2, \dots, \pm \frac{1}{2}(n - m). \quad (4)$$

Далее, *инварианты Казимира* \mathbf{C}_μ коммутируют со всеми генераторами \mathbf{X}_i алгебры \mathfrak{g} :

$$[\mathbf{C}_\mu, \mathbf{X}_i] = 0, \quad \forall \mu = 1 \rightarrow m, i = 1 \rightarrow n.$$

Число инвариантов (операторов) Казимира для данной алгебры Ли \mathfrak{g} определяется рангом этой алгебры.

Теорема 1. (Теорема Рака) Для каждой n -мерной алгебры Ли \mathfrak{g} ранга m существует в общей сложности m операторов Казимира \mathbf{C}_μ ($\mu = 1 \rightarrow m$), которые коммутируют с генераторами \mathbf{X}_i ($i = 1 \rightarrow n$) алгебры \mathfrak{g} .

Как следствие, все \mathbf{C}_μ также коммутируют с генераторами Картана \mathbf{H}_i :

$$[\mathbf{C}_\mu, \mathbf{H}_i] = 0, \quad \forall \mu, i = 1 \rightarrow m.$$

Таким образом, можно найти полный набор состояний, которые одновременно являются собственными состояниями всех \mathbf{C}_μ и \mathbf{H}_i . Определим кет-вектор следующего вида:

$$|c_1, c_2, \dots, c_m; h_1, h_2, \dots, h_m\rangle = |c_\mu; h_i\rangle,$$

где c_μ, h_i – собственные значения операторов $\mathbf{C}_\mu, \mathbf{H}_i$. Тогда

$$\mathbf{C}_\mu |c_\mu; h_i\rangle = c_\mu |c_\mu; h_i\rangle, \quad \mathbf{H}_i |c_\mu; h_i\rangle = h_i |c_\mu; h_i\rangle$$

для всех $\mu, i = 1 \rightarrow m$. Мы приходим к выводу, что инварианты Казимира \mathbf{C}_μ и образующие \mathbf{H}_i подалгебры Картана \mathfrak{K} позволяют нам помечать каждое состояние мультиплета, в то время как лестничные операторы \mathbf{E}_α позволяют перемещаться между состояниями внутри мультиплета, как показано на диаграмме Вейля. Таким образом, когда лестничный оператор \mathbf{E}_α воздействует на кет-вектор $|c_\mu; h_i\rangle$, он сдвигает собственное значение операторов \mathbf{H}_i на величину α_i в соответствии с

$$\mathbf{E}_\alpha |c_\mu; h_i\rangle \approx |c_\mu; h_i + \alpha_i\rangle.$$

3. Конформная группа $SO(4, 2)$

Специальная псевдоортогональная группа в шести измерениях, $SO(4, 2)$, соответствует группе вращений шестимерного псевдоевклидова пространства $\mathbb{R}^{4,2}$, или, что эквивалентно, множеству 6×6 ортогональных матриц, оставляющих квадратичную форму

$$Q(\mathbf{r}) = x_1^2 + x_2^2 + x_3^2 + x_4^2 - x_5^2 - x_6^2 = \mathbf{r}^T \mathbf{r},$$

где $\mathbf{r} = [x_1, x_2, x_3, x_4, x_5, x_6]^T$, инвариантной.

Структура соответствующей алгебры Ли $\mathfrak{so}(4, 2)$ определяется коммутационными свойствами её генераторов $\mathbf{L}_{\alpha\beta}$. $\mathbf{L}_{\alpha\beta}$ образуют базис алгебры $\mathfrak{so}(4, 2)$. Число независимых генераторов легко найти: из 36 возможных комбинаций индексов α и β шесть комбинаций исчезают в силу $\mathbf{L}_{\alpha\alpha} = 0$, это уменьшает число генераторов до 30. Более того, в силу $\mathbf{L}_{\alpha\beta} = -\mathbf{L}_{\beta\alpha}$ остаётся только 15 независимых генераторов, число которых также может быть получено с помощью формулы $n(n-1)/2$. Таким образом,

$$\mathbf{L} \Leftrightarrow \begin{bmatrix} 0 & \mathbf{L}_{12} & \mathbf{L}_{13} & \mathbf{L}_{14} & \mathbf{L}_{15} & \mathbf{L}_{16} \\ & 0 & \mathbf{L}_{23} & \mathbf{L}_{24} & \mathbf{L}_{25} & \mathbf{L}_{26} \\ & & 0 & \mathbf{L}_{34} & \mathbf{L}_{35} & \mathbf{L}_{36} \\ & & & 0 & \mathbf{L}_{45} & \mathbf{L}_{46} \\ & & & & 0 & \mathbf{L}_{56} \\ & & & & & 0 \end{bmatrix}. \quad (5)$$

Система пятнадцати генераторов $\mathbf{L}_{\alpha\beta}$ алгебры $\mathfrak{so}(4, 2)$ удовлетворяет следующим перестановочным соотношениям:

$$[\mathbf{L}_{\alpha\beta}, \mathbf{L}_{\gamma\delta}] = i (g_{\alpha\delta}\mathbf{L}_{\beta\gamma} + g_{\beta\gamma}\mathbf{L}_{\alpha\delta} - g_{\alpha\gamma}\mathbf{L}_{\beta\delta} - g_{\beta\delta}\mathbf{L}_{\alpha\gamma}), \quad (6)$$

где $\alpha, \beta, \gamma, \delta = 1, \dots, 6$, $\alpha \neq \beta$, $\gamma \neq \delta$, при этом $g_{11} = g_{22} = g_{33} = g_{44} = 1$, $g_{55} = g_{66} = -1$.

3.1. Водородная реализация алгебры $\mathfrak{so}(4, 2)$

В этом пункте рассмотрим *представление Барута* [9, 50] алгебры Ли $\mathfrak{so}(4, 2)$ конформной группы $SO(4, 2)$. Как отмечалось во введении, подгруппа $SO(4)$ имеет важное значение в полной спектр-генерирующей группе $SO(4, 2)$. Группа $SO(4)$ описывает вырождение уровней энергии атома водорода [7, 27]. В алгебре $\mathfrak{so}(4, 2)$ подалгебра $\mathfrak{so}(4)$ представлена следующими генераторами:

$$\begin{aligned} \mathbf{L}_1 &= \mathbf{L}_{23}, & \mathbf{L}_2 &= \mathbf{L}_{31}, & \mathbf{L}_3 &= \mathbf{L}_{12}, \\ \mathbf{A}_1 &= \mathbf{L}_{14}, & \mathbf{A}_2 &= \mathbf{L}_{24}, & \mathbf{A}_3 &= \mathbf{L}_{34}. \end{aligned} \quad (7)$$

Здесь $\mathbf{L}_1, \mathbf{L}_2, \mathbf{L}_3$ – генераторы углового момента, генераторы $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ соответствуют вектору Лапласа–Рунге–Ленца [27]. Коммутационные соотношения для (7) имеют вид

$$[\mathbf{L}_i, \mathbf{L}_j] = i\varepsilon_{ijk}\mathbf{L}_k, \quad [\mathbf{L}_i, \mathbf{A}_j] = i\varepsilon_{ijk}\mathbf{A}_k, \quad [\mathbf{A}_i, \mathbf{A}_j] = i\varepsilon_{ijk}\mathbf{L}_k.$$

Далее, полная спектр-генерирующая алгебра $\mathfrak{so}(4, 2)$ должна включать алгебру $\mathfrak{so}(2, 1)$, генераторы Δ_i ($i = 1 \rightarrow 3$) которой действуют на радиальную часть водородной волновой функции $\psi(r, \theta, \phi) = R_{n,l}(r)Y_l^m(\theta, \phi)$, а поскольку генераторы \mathbf{L}_i действуют только на угловую часть, то

$$[\mathbf{L}_i, \Delta_i] = 0.$$

Следовательно, генераторы Δ_i подалгебры $\mathfrak{so}(2, 1)$ не должны содержать общих индексов с генераторами \mathbf{L}_i углового момента. Таким образом,

$$\Delta_1 = \mathbf{L}_{46}, \quad \Delta_2 = \mathbf{L}_{45}, \quad \Delta_3 = \mathbf{L}_{56}.$$

Комбинируя генераторы \mathbf{A}_i с Δ_2 , получим элементы $\mathbf{L}_{15}, \mathbf{L}_{25}, \mathbf{L}_{35}$ пятого столбца в (5), которые обозначим символом \mathbf{B}_i :

$$[\Delta_2, \mathbf{A}_i] = i\mathbf{B}_i.$$

Три генератора \mathbf{B}_i являются компонентами вектора \mathbf{B} , который в некотором смысле сопряжён вектору Лапласа–Рунге–Ленца \mathbf{A} .

Аналогично, элементы $\mathbf{L}_{16}, \mathbf{L}_{26}, \mathbf{L}_{36}$ шестого столбца в (5) получаются посредством коммутации вектора \mathbf{A} с Δ_1 :

$$[\Delta_1, \mathbf{A}_i] = i\mathbf{G}_i.$$

В свою очередь, три генератора Γ_i являются компонентами вектора, обозначенного Барутом Γ .

Таким образом, все 15 генераторов алгебры $\mathfrak{so}(4, 2)$ могут быть представлены в матричной форме

$$\mathbf{L} \Leftrightarrow \begin{bmatrix} 0 & L_3 & -L_2 & A_1 & B_1 & \Gamma_1 \\ & 0 & L_1 & A_2 & B_2 & \Gamma_2 \\ & & 0 & A_3 & B_3 & \Gamma_3 \\ & & & 0 & \Delta_2 & \Delta_1 \\ & & & & 0 & \Delta_3 \\ & & & & & 0 \end{bmatrix}.$$

3.2. Подалгебра Картана

Найдём максимальное подмножество коммутирующих генераторов алгебры $\mathfrak{so}(4, 2)$. Как известно, два генератора коммутируют, если они не имеют общих индексов. Легко видеть, что среди генераторов алгебры $\mathfrak{so}(4, 2)$ этому условию удовлетворяют три генератора \mathbf{L}_{12} , \mathbf{L}_{34} и \mathbf{L}_{56} (т. е. L_3 , A_3 и Δ_3 соответственно):

$$[L_3, A_3] = [L_3, \Delta_3] = [A_3, \Delta_3] = 0.$$

Триплет $\{L_3, A_3, \Delta_3\}$ образует базис максимальной абелевой подалгебры $\mathfrak{K} \subset \mathfrak{so}(4, 2)$ (подалгебры Картана). L_3 , A_3 и Δ_3 называются генераторами Картана. Размерность подалгебры \mathfrak{K} определяет ранг алгебры Ли $\mathfrak{so}(4, 2)$. Как следствие, все корневые и весовые диаграммы для $\mathfrak{so}(4, 2)$ будут трёхмерными.

Инварианты Казимира. В силу теоремы Рака и того факта, что алгебра $\mathfrak{so}(4, 2)$ имеет ранг 3, можно ожидать, что группа $SO(4, 2)$ допускает три независимых инварианта Казимира \mathbf{C}_μ ($\mu = 1 \rightarrow 3$), которые коммутируют со всеми генераторами алгебры $\mathfrak{so}(4, 2)$. Наиболее важным оператором Казимира группы $SO(4, 2)$ является квадратичная комбинация инвариантов различных подгрупп:

$$\mathbf{C}_2 = \mathbf{L}^2 + \mathbf{A}^2 - \mathbf{B}^2 - \mathbf{\Gamma}^2 + \Delta_3^2 - \Delta_1^2 - \Delta_2^2.$$

Здесь $\mathbf{L}^2 + \mathbf{A}^2$ и $\Delta_3^2 - \Delta_1^2 - \Delta_2^2$ известны как операторы Казимира групп $SO(4)$ и $SO(2, 1)$ соответственно. Остальные два оператора Казимира группы $SO(4, 2)$ являются полиномами третьей и четвёртой степени относительно генераторов алгебры $\mathfrak{so}(4, 2)$:

$$\mathbf{C}_3 = \frac{1}{48} \varepsilon_{abcdef} \mathbf{L}^{ab} \mathbf{L}^{cd} \mathbf{L}^{ef}, \quad \mathbf{C}_4 = \mathbf{L}_{ab} \mathbf{L}^{bc} \mathbf{L}_{cd} \mathbf{L}^{da}.$$

4. Группа $SU(2, 2)$ и базис Яо

Двухлистное накрытие конформной группы $SO(4, 2)$ изоморфно группе $SU(2, 2)$. $SU(2, 2)$ (группа псевдоунитарных унимодулярных 4×4 матриц) определяется как группа преобразований четырёхмерного комплексного пространства \mathbb{C}^4 (твисторного пространства), оставляющей инвариантной квадратичную форму $|Z_1|^2 + |Z_2|^2 -$

$|Z_3|^2 - |Z_4|^2$. При этом *твисторы* (векторы пространства \mathbb{C}^4) определяются как редуцированные спиноры для конформной группы (см. Приложение Б в [51]).

Базис алгебры $\mathfrak{su}(2, 2)$ относительно максимальной компактной подгруппы

$$K = \text{SU}(2) \otimes \text{SU}(2) \otimes \text{U}(1)$$

связан с генераторами алгебры $\mathfrak{so}(4, 2)$ следующими соотношениями [47–49]:

$$\mathbf{K}_1 = 1/2 (\mathbf{L}_{23} + \mathbf{L}_{14}), \quad \mathbf{K}_2 = 1/2 (\mathbf{L}_{31} + \mathbf{L}_{24}), \quad \mathbf{K}_3 = 1/2 (\mathbf{L}_{12} + \mathbf{L}_{34}), \quad (8)$$

$$\mathbf{J}_1 = 1/2 (\mathbf{L}_{23} - \mathbf{L}_{14}), \quad \mathbf{J}_2 = 1/2 (\mathbf{L}_{31} - \mathbf{L}_{24}), \quad \mathbf{J}_3 = 1/2 (\mathbf{L}_{12} - \mathbf{L}_{34}), \quad (9)$$

$$\mathbf{T}_1 = 1/2 (-\mathbf{L}_{15} - \mathbf{L}_{26}), \quad \mathbf{T}_2 = 1/2 (\mathbf{L}_{25} - \mathbf{L}_{16}), \quad \mathbf{T}_0 = 1/2 (-\mathbf{L}_{12} - \mathbf{L}_{56}), \quad (10)$$

$$\mathbf{S}_1 = 1/2 (-\mathbf{L}_{15} + \mathbf{L}_{26}), \quad \mathbf{S}_2 = 1/2 (-\mathbf{L}_{25} - \mathbf{L}_{16}), \quad \mathbf{S}_0 = 1/2 (\mathbf{L}_{12} - \mathbf{L}_{56}), \quad (11)$$

$$\mathbf{P}_1 = 1/2 (-\mathbf{L}_{35} - \mathbf{L}_{46}), \quad \mathbf{P}_2 = 1/2 (\mathbf{L}_{45} - \mathbf{L}_{36}), \quad \mathbf{P}_0 = 1/2 (-\mathbf{L}_{34} - \mathbf{L}_{56}), \quad (12)$$

$$\mathbf{Q}_1 = 1/2 (\mathbf{L}_{35} - \mathbf{L}_{46}), \quad \mathbf{Q}_2 = 1/2 (\mathbf{L}_{45} + \mathbf{L}_{36}), \quad \mathbf{Q}_0 = 1/2 (\mathbf{L}_{34} - \mathbf{L}_{56}). \quad (13)$$

Базис Яо (8)–(13) содержит 18 генераторов, что создаёт избыточную систему генераторов для алгебры $\mathfrak{so}(4, 2)$, поскольку последняя состоит из 15 независимых генераторов. Базис алгебры $\mathfrak{so}(4, 2)$ можно получить из (8)–(13), исключив три генератора с помощью соотношений

$$\mathbf{J}_3 + \mathbf{K}_3 = \mathbf{S}_0 - \mathbf{T}_0 = \mathbf{L}_{12} = \mathbf{L}_3, \quad (14)$$

$$\mathbf{J}_3 - \mathbf{K}_3 = \mathbf{P}_0 - \mathbf{Q}_0 = -\mathbf{L}_{34} = -\mathbf{A}_3, \quad (15)$$

$$\mathbf{P}_0 + \mathbf{Q}_0 = \mathbf{S}_0 + \mathbf{T}_0 = -\mathbf{L}_{56} = -\mathbf{\Delta}_3. \quad (16)$$

Легко видеть, что шесть генераторов $\mathbf{K}_3, \mathbf{J}_3, \mathbf{P}_0, \mathbf{Q}_0, \mathbf{S}_0, \mathbf{T}_0$ базиса Яо в силу соотношений (14)–(16) эмулируют подалгебру Картана $\mathfrak{K} = \{\mathbf{L}_3, \mathbf{A}_3, \mathbf{\Delta}_3\}$ алгебры Ли $\mathfrak{so}(4, 2)$.

Из оставшихся 12 генераторов базиса (8)–(13) образуем генераторы Вейля посредством следующих линейных комбинаций

$$\mathbf{J}_\pm = \mathbf{J}_1 \pm i\mathbf{J}_2, \quad \mathbf{P}_\pm = \mathbf{P}_1 \pm i\mathbf{P}_2, \quad \mathbf{S}_\pm = \mathbf{S}_1 \pm i\mathbf{S}_2,$$

$$\mathbf{K}_\pm = \mathbf{K}_1 \pm i\mathbf{K}_2, \quad \mathbf{Q}_\pm = \mathbf{Q}_1 \pm i\mathbf{Q}_2, \quad \mathbf{T}_\pm = \mathbf{T}_1 \pm i\mathbf{T}_2. \quad (17)$$

Изучение явного вида и корневой структуры генераторов Вейля в зависимости от различных подалгебр алгебры $\mathfrak{so}(4, 2)$ выходит за ограниченные рамки данной статьи и будет рассмотрено в следующей работе.

Литература

1. Scerri E. Periodic Table: its story and its significance. New York: Oxford University Press, 2019.
2. Nakala R. The periodic law in mathematical form // J. Phys. Chem. 1952. Vol. 56. P. 178–181.
3. Румер Ю.Б., Фет А.И. Группа Spin(4) и таблица Менделеева // ТМФ. 1971. Т. 9. С. 203–209.
4. Маделунг Э. Математический аппарат физики. М.: Наука, 1968.
5. Ostrovsky V.N. What and how physics contributes to understanding the periodic law // Foundations of Chemistry. 2001. Vol. 3. P. 145–182.
6. Karapetoff V. A chart of consecutive sets of electronic orbits within atoms of chemical elements // J. Franklin Inst. 1930. Vol. 210. P. 609–614.
7. Фок В.А. Атом водорода и неевклидова геометрия // Изв. АН СССР. Сер. VII. Отделение мат. и естеств. наук. 1935. № 2. С. 169–179.
8. Конопельченко Б.Г. Группа $SO(2, 4) + R$ и таблица Менделеева: препринт. Новосибирск: СО РАН СССР, Ин-т ядерной физики, 1972.
9. Barut A.O. Group Structure of the Periodic System // The Structure of Matter: Rutherford Centennial Symposium / Ed. by B.G. Wybourne. Christchurch, New Zeland: University of Canterbury Press, 1972. P. 126–136.
10. Ostrovsky V.N. Group theory and periodic system of elements // AIP Conference Proceedings. 1996. Vol. 365. P. 191–216.
11. Wulfman C.E. Dynamical Groups in Atomic and Molecular Physics // Recent Advances in Group Theory and Their Application to Spectroscopy / Ed. by J.C. Donini. New York: Plenum Press, 1978. P. 329–403.
12. Варламов В.В. Алгебраическая квантовая механика. I: Основные определения // Математические структуры и моделирование. 2020. № 2 (54). С. 4–23.
13. Варламов В.В. Алгебраическая квантовая механика. II: S-матрица // Математические структуры и моделирование. 2021. № 1 (57). С. 3–24.
14. Варламов В.В. Алгебраическая квантовая механика. III: Вопросы интерпретации // Математические структуры и моделирование. 2021. № 3 (59). С. 4–26.
15. Варламов В.В. Алгебраическое квантование и спинорная структура // Математические структуры и моделирование. 2022. № 1 (61). С. 5–25.
16. Novaro O., Berrondo M. Approximate symmetry of the periodic table // J. Phys. B: Atom. Molec. Phys. 1972. Vol. 5. P. 1104–1110.
17. Berrondo M., Novaro O. On a geometrical realization of the Aufbau scheme // J. Phys. B: Atom. Molec. Phys. 1973. Vol. 6. P. 761–769.
18. Löwdin P.-O. Some Comments on the Periodic System of the Elements // Int. J. Quant. Chem. 1969. Vol. 3. P. 331–334.
19. Thyssen P., Ceulemans A. Shattered Symmetry: Group Theory from the Eightfold Way to the Periodic Table. New York: Oxford University Press, 2017.
20. Кораблева Т.П., Корольков Д.В. Теория периодической системы. СПб.: Изд-во С.-Петербург. ун-та, 2005.
21. Scerri E. The Electronic Configuration Model, Quantum Mechanics and Reduction // Brit. J. Phil. Sci. 1991. Vol. 42. P. 309–325.
22. Lombardi O., Labarca M. The ontological autonomy of the chemical world // Foundations of Chemistry. 2005. Vol. 7. P. 125–148.

23. Sanderson R.T. An Explanation of Chemical Variations within Periodic Major Groups // J. Amer. Chem. Soc. 1952. Vol. 74, Iss. 19. P. 4792–4794.
24. Neubert D. Double Shell Structure of the Periodic System of the Elements // Zeitschrift für Naturforschung. 1970. Vol. 25a. P. 210–217.
25. Odabasi H. Some Evidence about the Dynamical Group $SO(4,2)$. Symmetries of the Periodic Table of Elements // Int. J. Quant. Chem. 1973. Vol. 7, Suppl. 7. P. 23–33.
26. Novaro O. Group Theoretical Aspects of the Periodic Table of the Elements // Journal of Molecular Structure. 1989. Vol. 199. P. 103–118.
27. Pauli W. Über das Wasserstoffspektrum vom Standpunkt der neuen Quantenmechanik // Z. Phys. 1926. Vol. 36. P. 336–363.
28. Fock V. Zur Theorie des Wasserstoffatoms // Z. Phys. 1935. Vol. 98.3. P. 145–154.
29. Bargmann V. Zur Theorie des Wasserstoffatoms // Z. Phys. 1936. Vol. 99.7. P. 576–582.
30. Fet A.I. The System of Elements from the Group-Theoretic Viewpoint. Новосибирск, 1979. 45 с. (Препр. / АН СССР. Сиб. отд-ние ИНХ СО РАН СССР; № 1).
31. Fet A.I. The Madelung Numbers and the System of Chemical Elements // Теоретико-групповые методы в физике: тр. междунар. семинара, Звенигород, 1979. М.: Наука, 1980. Т. 1. С. 327.
32. Ostrovsky V.N. Dynamic symmetry of atomic potential // J. Phys. B: At. Mol. Phys. 1981. Vol. 14. P. 4425–4439.
33. Fet A.I. The System of Elements from the Group-Theoretic Viewpoint // R. Hefferlin. Periodic Systems and their Relations to the Systematic Analysis of Molecular Data. Lewiston; N.Y.: Edwin Mellen Press, 1989. P. 41–86.
34. Фет А.И. Группа симметрии химических элементов. Новосибирск: Наука, 2010.
35. Варламов В.В. Теоретико-групповое описание периодической системы элементов // Математические структуры и моделирование. 2018. № 2 (46). С. 5–23.
36. Варламов В.В. Теоретико-групповое описание периодической системы элементов. II: Таблица Сиборга // Математические структуры и моделирование. 2019. № 1 (49). С. 5–21.
37. Варламов В.В. Теоретико-групповое описание периодической системы элементов. III: 10-периодическое расширение // Математические структуры и моделирование. 2019. № 3 (51). С. 5–20.
38. Varlamov V.V., Pavlova L.D, Babushkina O.S. Group Theoretical Description of the Periodic System // Symmetry. 2022. Vol. 14. Art. 137.
39. Дынкин Е.Б. Классификация простых групп Ли // Матем. сб. 1946. Т. 60. С. 347–352.
40. Mazurs E.G. Graphic Representations of the Periodic System During One Hundred Years. Tuscaloosa: University of Alabama Press, 1974.
41. van Spronsen J.W. The Periodic System of Chemical Elements. Amsterdam; New York: Elsevier, 1969.
42. Varlamov V.V. Universal Coverings of Orthogonal Groups // Adv. Appl. Clifford Algebras. 2004. Vol. 14. P. 81–168.
43. Thomas L.H. On unitary representations of the group of de Sitter space // Ann. Math. 1941. Vol. 42. P. 113–126.
44. Murai Y. On the Group of Transformations in Six-dimensional Space // Progr. Theor. Phys. 1953. Vol. 9. P. 147–168.
45. Murai Y. On the Group of Transformations in Six-dimensional Space, II // Progr. Theor. Phys. 1954. Vol. 11. P. 441–448.

46. Kihlberg A., Müller V.F., Halbwachs F. Unitary Irreducible Representations of $SU(2,2)$ // Commun. Math. Phys. 1966. Vol. 3. P. 194–217.
47. Yao T. Unitary Irreducible Representations of $SU(2,2)$. I // J. Math. Phys. 1967. Vol. 8. P. 1931–1954.
48. Yao T. Unitary Irreducible Representations of $SU(2,2)$. II // J. Math. Phys. 1968. Vol. 9. P. 1615–1626.
49. Yao T. Unitary Irreducible Representations of $SU(2,2)$. III. Reduction with Respect to an IsoPoincaré Subgroup // J. Math. Phys. 1971. Vol. 12. P. 315–342.
50. Adams B.G., Čížek, Paldus J. Representation Theory of $SO(4,2)$ for the Perturbation Treatment of Hydrogenic-Type Hamiltonians by Algebraic Methods // Int. J. Quant. Chem. 1982. Vol. 21. P. 153–171.
51. Варламов В.В. О системе аксиом нелокальной квантовой теории // Математические структуры и моделирование. 2017. № 4 (44). С. 5–25.

GROUP THEORETICAL DESCRIPTION OF PERIODIC SYSTEM OF ELEMENTS. IV: GROUP ALGEBRA

V.V. Varlamov

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

Siberian State Industrial University, Novokuznetsk, Russia

Abstract. The structure of the group algebra of a conformal group (the group underlying the group-theoretic description of the periodic system of chemical elements) is considered within the framework of a twofold covering. The hydrogen realization of the Cartan subalgebra and Weyl generators of the group algebra is studied.

Keywords: periodic law, Mendeleev table, conformal group, group algebra, Cartan subalgebra, Weyl generators.

Дата поступления в редакцию: 29.11.2023

ФИЗИЧЕСКОЕ СОСТОЯНИЕ ОБЪЕКТОВ ПРОШЛОГО И БУДУЩЕГО

А.К. Гуц

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

Международный инновационный университет, Сочи, Россия

Аннотация. Теория абсолютного пространства-времени, на которой основана теория относительности, благодаря Минковскому, говорит о том, что события прошлого, настоящего и будущего равноправны относительно их бытия, т. е. они всегда существуют и никуда не исчезают. В статье рассматривается вопрос о физическом состоянии их существования. Строится формализм в рамках теории относительности, согласно которому поскольку мы не видим и не взаимодействуем с объектами прошлого и будущего, то их состояния представляют собой так называемую призрачную материю, характеризующуюся нулевым тензором энергии-импульса. С течением координатного (глобального) времени призрачная материя переходит в настоящее и обнаруживается как реальность, а затем вновь переходит в призрачную и становится объектом прошлого.

Ключевые слова: теория абсолютного пространства-времени, прошлое, будущее, призрачная материя.

Введение

Почему мы не видим своего прошлого? Мы же наблюдаем прошлое звёзд и галактик, и только их прошлое, поскольку свет от них идёт годы, и, соответственно, мы видим их такими, какими они были миллионы лет назад.

Я в прошлом – это такое же событие в пространстве-времени, как и события в жизни звёзд и галактик. Это точки в 4-мерном Мире событий, или в пространстве-времени.

Объяснение простое – всё дело в скорости света, во-первых, а во-вторых, в работе нашего мозга.

Действительно, если моя координата $(x, 0, 0)$ в пространстве в прошлом, отстоящем на T сек от настоящего момента x^0/c , то $x = cT$. Следовательно, если T секунд назад я был на расстоянии $x = 10$ м, то $T = 3 \cdot 10^{-7}$ сек. Вряд ли мозг в состоянии показать нам себя за столько секунд назад. А если $x = 30000$ км, то $T = 1/10$ сек. Опять-таки, сомнительно, что мозг покажет нам, как мы выглядели 0,1 секунды назад на столь огромном расстоянии. Даже если допустить, что мы умеем перемещаться почти мгновенно в пространстве, а мозг показывать нас, то показывает он нам какую-то точку, и вопрос, сумеем ли мы различить себя в той точке?

Можем ли мы из прошлого влиять на настоящее? Формально, да, можем. Выстреливший в нас, стреляет из прошлого; излучение звезды, т. е. излучение из прошлого, может нанести урон на Земле.

Поэтому естественно предположить, что я вчерашний могу воздействовать на себя сегодняшнего. Вопрос, в чём и как это проявляется? Например, в воспоминаниях. Наши воспоминания – это сигналы, информация, идущая из прошлого. Мы же этот факт традиционно объясняем как некоторый процесс в мозгу, который запускается в тот момент, когда мы что-то пытаемся вспомнить. Однако можно предположить, что, пытаясь что-то вспомнить, мы обращаемся к точкам-событиям, лежащим в Прошлом. Другими словами, мы имеем дело с темпоральным взаимодействием «прошлое–настоящее», которое является *квантовым запутыванием во времени* [1–3], если учитывать временной масштаб отстояния событий прошлого и настоящего, состоящий в передаче образов самого себя из Прошлого, о котором говорилось выше.

С каким типом квантового взаимодействия в таком случае мы имеем дела? С *квантовой корреляцией во времени* [1, 2].

Вспомним, что события в Мире событий, в пространстве-времени, *заполнены субстанцией*. Об этом сразу писал Минковский [4], об этом напомнил и обстоятельно пояснил «что и почему» А.Д. Александров [4]. Применительно к нашей задаче, говоря о квантовой корреляции во времени между событиями прошлого и настоящего, следует разобраться с физическим состоянием этой субстанции, этих физических объектов (вещей).

Состояния объектов настоящего – это то, что описывается ненулевым тензором энергии-импульса. Они обладают энергией, они находятся в движении (импульс). А каково физическое состояние объектов Прошлого? Ответ очевиден, мы их не ощущаем, классически не взаимодействуем (не толкаем, не мнём, не рвём и т. д.), т. е. их тензор энергии-импульса нулевой.

1. Призрачные частицы и призрачная материя

Частица-призрак – это частица, тензор момента-импульса которой является нулевым, но при этом ненулевым является её ток, который в случае биспинора равен

$$j^i = \psi^+ \gamma^i \psi.$$

Следовательно, такая частица не несёт ни энергии, ни импульса. И в этом смысле она не является материальной, реальной.

Впервые такие призрачные неабелевы решения (уравнений Янга–Миллса) были найдены в 1967 г. [6]. Напомним, что кванты полей Янга–Миллса являются векторными частицами (т. е. бозонами со спином 1) и обладают нулевой массой. Однако с помощью механизма спонтанного нарушения симметрии физические поля Янга–Миллса могут приобретать ненулевую массу. При рассмотрении слабого взаимодействия квантом поля Янга–Миллса считают W-частицу, имеющую заряд +1, 0 или –1. Для сильного взаимодействия квантом поля Янга–Миллса являются глюоны, удерживающие вместе протоны и нейтроны. Фактически призрачность квантов Янга–Миллса говорит о призрачности глюонов. А это порождает ожидание о существо-

вании призрачных протонов и нейтронов. Или, поскольку и те и другие являются спинорными частицами, т. е. частицами со спином $1/2$, то достаточно убедиться в существовании *призрачных спинорных частиц*.

Распространяющиеся во внешнем пространстве частицы-призраки были открыты в 1972 г. Гриффитсом [7], а затем соответствующие решения уравнений Эйнштейна–Дирака были найдены и опубликованы в многочисленных работах других авторов в 1970-1980-х гг., а также в XXI в. Подробный обзор дан в [8].

«Первой реакцией, – писал в 1974 г. М. Новелло [9], – на эти решения может быть отказ от них на физических основаниях». Тем не менее, он высказал убеждение, что призрачные частицы должны быть изучены более внимательно. Сам он обнаружил удивительный результат: привёл пример пространства-времени, порождённого нейтрино, в котором само это нейтрино есть линейная комбинация нейтрино-призраков.

Однако до работы Е.В. Палешевой [10, 11] никто так и не смог найти хоть какое-нибудь место в картине Мироздания для теоретически обнаруженных частиц-призраков. Она отождествила частицы-призраки с теньевыми частицами Дойча [12], который считал, что это частицы, принадлежащие параллельной вселенной Эверетта.

Объединяя результаты Палешевой и Новелло, можно высказать следующую гипотезу.

Вселенная, наше присутствие в которой мы осознаём, состоит из *реальных* частиц, т. е. частиц с ненулевым тензором энергии-импульса. Частицы-призраки – это гости из параллельных вселенных Эверетта; они объекты параллельных вселенных.

Но параллельных вселенных бесконечно много; все они симметричны относительно нашего анализа (нет выделенной «нашей» Вселенной). Следовательно, казалось бы, могут существовать только частицы-призраки. Предполагаем, что энергия и импульс придаются частице в конкретной рассматриваемой, т. е. зафиксированной чьим-то сознанием, вселенной, если, с точки зрения математики, она есть линейная комбинация частиц-призраков. Но для разложения частицы в линейную комбинацию требуется некий механизм, присутствующий во вселенной, который осуществляет и подтверждает факт разложения. Очевидно, что это тот же механизм, который фиксировал конкретную вселенную. И механизм этот есть сознание, есть наблюдатель, присутствующий, живущий в этой вселенной.

Сознание своим вниманием к окружающей его вселенной совершает акт творения, выражающийся в порождении линейных комбинаций частиц-призраков: от хаоса частиц-призраков к порядку=«линейным комбинациям», от простого к сложному.

Таким образом, «наши» частицы – это те, что обладают энергией и импульсом, т. е. для них $T_{ik} \neq 0$. Они *заполняют* наше пространство-время. Частицы призраки *заполняют* параллельные миры Эверетта. Пустого, т. е. незаполненного, пространства-времени быть не может [4, 5]. Однако существуют частицы, которые путешествуют из нашего пространства-времени в параллельное и обратно; они то реальные, то призрачные. Об этом говорится ниже в § 3, 4.

Что особо важно отметить: Елена Палешева показала, что частицы-призраки, т. е. частицы параллельной вселенной, взаимодействуют с нашими частицами и это

взаимодействие проявляется в форме квантовой интерференции.

Вспомним, что нам известно и иное квантовое взаимодействие частиц нашей Вселенной и параллельной. Это квантовая сцепленность (запутанность). В статье [8] было показано, что данный тип *несилового* взаимодействия возможен для частиц разных вселенных, т. е. квантовая запутанность может иметь место между реальной и призрачной частицами.

Но нас в данной статье интересует вопрос о переходе реальной частицы в призрачную и наоборот.

2. Примеры частиц-призраков

Рассмотрим спинорные частицы в пространстве-времени Минковского, описываемые уравнением Дирака:

$$i\hbar\gamma^{(k)}\frac{\partial\psi}{\partial x^k} - mc\psi = 0. \quad (1)$$

$$x = (x^0, x^1, x^2, x^3).$$

Тогда биспинор

$$\psi(x) = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix} e^{\frac{mc}{\hbar}x^2 + f(x^0 + x^3) + ig(x^0 + x^3)} \quad (2)$$

является решением уравнения Дирака. Здесь $g(x^0 + x^3)$ и $f(x^0 + x^3)$ – гладкие вещественные функции.

Опираясь на результаты теоремы, приведённой ниже в § 5, получаем, что (2) описывает спинорного призрака только лишь в том случае, когда $g(x^0 + x) = const \in \mathbb{R}$.

В таком случае, примером реальной частицы является биспинор:

$$\psi(x) = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix} e^{\frac{mc}{\hbar}x^2 + i(x^0 + x^3)}, \quad (3)$$

а для спинорного призрака имеем

$$\phi(x) = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix} e^{\frac{mc}{\hbar}x^2}. \quad (4)$$

Отметим важное обстоятельство: биспиноры с $T_{ik} \neq 0$ во всём пространстве-времени – это субстанция, заполняющая его и фиксирующая его абсолютность, его

цельность (геодезическую полноту), его вечность в соответствии с постулатом абсолютного Мира событий Минковского. Напротив, те биспиноры-призраки, для которых $T_{ik} \equiv 0$ во всём пространстве-времени, – это субстанция, заполняющая некоторое параллельное пространство-время. Но параллельных пространств-времён бесконечно много, поэтому уместен вопрос: какое именно параллельное пространство-время они фиксируют, выделяют как нечто целое (геодезически полное), заполняют и т. д.? На уровне классической логики трудно ответить на этот вопрос, но множественность этих миров легко описывается в неклассической логике и демонстрируется в § 6.

3. Переход реальной частицы в призрачную с течением времени в момент $x^0 = a$

Рассмотрим биспинор

$$\psi(x) = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix} e^{\frac{mc}{\hbar}x^2 + i(x^0 + x^3)\theta(a - x^0)}, \quad (5)$$

где

$$\theta(\tau) = \begin{cases} 1, & \tau \geq 0, \\ 0, & \tau < 0. \end{cases} \quad (6)$$

Как видим, биспинор (5) описывает частицу, которая реальна с момента $x^0 \geq a$, но призрачна в прошлом момента a , т. е. при $x^0 < a$.

Представим теперь наличие течения внешнего времени x^5 и рассмотрим биспинор

$$\psi(x) = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix} e^{\frac{mc}{\hbar}x^2 + i(x^0 + x^3)\theta(x^5 - x^0)}. \quad (7)$$

Течение внешнего времени x^5 отодвигает верхнюю границу времени $x^0 = x^5$ 4-мерного реального бытия частицы (7) в будущее, и то, что ещё было реальным, становится призрачным.

Иначе говоря, частицы (7), уходя в прошлое, становятся призрачными. Поскольку в § 1 мы связывали реальность с сознанием людей, то частицы (7) становятся с течением времени недоступными для их осознания людьми – они составляют то Прошлое, которое *ушло* и стало *недоступным*. Но они составляют параллельную вселенную и доступны сознанию, присутствующему в этой вселенной.

Введение внешнего времени x^5 означает переход к 5-мерному Гиперпространству, которое впервые появилось в работах Калуцы (1919), Клейна (1926) и совет-

ского физика Г.А. Мандела (1926), о чём сейчас подзабыли [13]. Геометрия Гиперпространства описывается, например, метрикой:

$$dI^2 = dx^{5^2} + dx^{0^2} - \sum_{\alpha=1}^3 dx^\alpha dx^\beta.$$

4. Настоящее – миг между прошлым и будущим

Вместо функции (6) рассмотрим функцию

$$\theta_\tau(t) = \begin{cases} 0, & t > \tau, \\ 1, & 0 \leq t \leq \tau, \\ 0, & t < 0, \end{cases} \quad (8)$$

где $\tau > 0$ – фиксированная фундаментальная константа, значение которой характеризует сущность бытия материального мира.

Рассмотрим теперь биспинор

$$\psi_\tau(x) = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix} e^{\frac{mc}{\hbar}x^2 + i(x^0 + x^3)\theta_\tau(x^5 - x^0)}. \quad (9)$$

Физическое состояние частицы, которую он описывает, имеет три значения:

$$\theta_\tau(t) = \begin{cases} \text{призрак,} & x^0 < x^5, \\ \text{реальность,} & x^5 \leq x^0 \leq x^5 + \tau, \\ \text{призрак,} & x^0 > x^5 + \tau. \end{cases} \quad (10)$$

Следовательно, с течением внешнего времени x^5 в пространстве-времени движется в будущее во времени x^0 узкая полоса реальности частицы, описываемой биспинором (9) шириной τ , которая в будущем, т. е. выше данной полосы, и в прошлом, т. е. ниже данной полосы, является призраком.

Другими словами, наше будущее и наше прошлое – это призрачное бытие частиц. И только в узкой полосе времени x^0 составляющие нас атомы находятся в реальном состоянии.

Не противоречит ли этот пример тому, что мы выше оговаривали как целостность (геодезическую полноту) нашей (и прочих) вселенной, нашего пространства-времени? Пример как бы говорит о *склейке* разных параллельных пространств-времён, да ещё меняющейся во внешнем времени x^5 .

Кстати, о похожего рода склейках писал в книге «Неоднозначное мироздание» Ю.А. Лебедев [14], а математически описывал автор этой статьи в [11]. Однако тщательной продуманной математической модели склеек не существует.

Ответить на заданный вопрос достаточно сложно... поскольку мы увязывали непризрачность с сознанием. В таком случае можно приведённый пример с переходящим состоянием частиц «реальность \Leftrightarrow призрак» объяснить как результат переходящего осознания конкретного поколения, а целостность вселенной – как результат совокупного осознания человечества. В первом случае имеем *конкретное бытие* частицы по Гегелю, а во-втором – *чистое бытие* вселенной, т. е. *ничто* [15, с. 139–140].

Чистое бытие или ничто, чистое ничто, – есть, по сути, пустое созерцание (мышление) [15, с. 139–140]. И это схоже с тем, как Минковский заполнял события субстанцией, используя это слово, чтобы не говорить, чем именно они заполняются – материей или излучением, т. е. отказываясь от конкретики. Иначе говоря, пространство-время *есть*, оно мыслится, и этого достаточно. Абсолютность Мира событий, пространства-времени, всего лишь обозначается, как обозначается ничто, но не конкретизируется. Тем самым констатируется, что пространство-время *не совсем* абсолютно, как об этом писал А.Д. Александров [16, с. 103]. Слова «не совсем» превращают высказывание об абсолютности Мира событий из высказывания однозначно истинного (или ложного), высказывания классической логики, в высказывание модальной логики, истинность которого многозначна, и это решает парадокс дедушки, открывая дорогу конструированию (квантовой) машины времени [5].

5. Корректный выбор биспинора (9)

Биспинор (9) имеет недостаток – функция $\theta_r(x)$ разрывна, и точки разрыва совпадают со сменами настоящего, прошлого и будущего. Такая ситуация не может быть признана корректной, поскольку при дифференцировании биспинора при подстановке в уравнение Дирака возникают δ -функции.

Покажем, что данная ситуация вполне исправима, и биспинор может быть выбран вполне гладким, и, следовательно, переходы «реальность \Leftrightarrow призрак» происходят без каких-либо особенностей.

Это означает, что данные переходы являются естественными так же, как естественен для нас уход настоящего в прошлое. Реальная материя, состоящая из реальных частиц, превращается в призрачную и в классически недоступную материю, ушедшую в Прошлое. И так же естественно призрачная материя Будущего превращается в реальную материю Настоящего.

Чтобы получить нужный биспинор, воспользуемся теоремой, доказанной Е.В. Палешевой ([11, с. 309], [17])

Теорема 1. Пусть $\psi = u \cdot G(x)$ – решение уравнения Дирака, при этом полагаем $\psi^* \psi \neq 0$,

$$G(x) = f(x) + i \cdot g(x),$$

где $f(x)$ и $g(x)$ – гладкие вещественные функции, а

$$u = \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \end{bmatrix},$$

где $\forall i \quad u_i \in \mathbb{C}$. В рассмотренных условиях ψ является спинорным духом $\iff g(x) = a \cdot f(x)$, где $a = \text{const} \in \mathbb{R}$.

Нетрудно видеть, что в соответствии с этой теоремой, если взять функцию $g(x) = h(x)\phi_\tau(x^5 - x^0)$, где функции h, ϕ_τ гладкие и функция $\phi_\tau(t) = 0$ вне интервала $(0, \tau)$, а внутри отлична от нуля, то получаем биспинор, который претерпевает с течением времени x^0 те же метаморфозы «реальность \rightleftharpoons призрак», что и рассмотренный биспинор $\psi_\tau(x)$ из (9).

Обратим внимание на то, что полученный биспинор удовлетворяет уравнению Дирака (1).

6. Множественность параллельных вселенных

Частицы-призраки, или теньевые частицы Дойча, – это объекты параллельных вселенных. Однако до сих пор мы не видели каких-либо доводов в пользу их множественности.

Всё дело в двузначной логике, которой мы неявно пользуемся, рассуждая и излагая получаемые результаты. У нас есть обычные частицы, реальные, «наши», обладающие ненулевым тензором энергии-импульса, и призраки, для которых $T_{ik} \equiv 0$.

Мы перейдём к многозначной логике, и у нас появится множество различных параллельных вселенных, если вместо $T_{ik} \equiv 0$ мы примем как условие прозрачности требование

$$T_{ik} = \varepsilon, \quad \varepsilon \neq 0, \quad \varepsilon^2 = 0. \tag{11}$$

В классической теории действительных чисел такого числа ε быть не может, но в неклассическом инфинитозимальном анализе они существуют. Мы расширяем поле действительных чисел \mathbb{R} до кольца $R = \mathbb{R} \cup D$, где $D = \{d : d^2 = 0\}$.

А такой теории существуют производные, интегралы, т. е. возможен гладкий неклассический инфинитозимальный анализ Кока–Ловера [18].

Для интерпретации такого дифференциального исчисления Кока–Ловера не подходит теория множеств Кантора, а используются гладкие топосы [19] – объекты, похожие на множества, но опирающиеся на многозначную интуиционистскую логику.

При интерпретации построенной теории в топосах и появляется бесчисленное многообразие различных вселенных, в которых обитают частицы-призраки, а также *наша* Вселенная с реальными частицами.

Можно обнаружить множественность вселенных, не обращаясь к условию (11). Достаточно положить, что в уравнении Дирака (1) скорость света $c = c_0 + d$, $c_0 \in$

\mathbb{R} , $d^2 = 0$. Тогда мы получим множество вселенных, каждая из которых имеет различную скорость света.

Аналогичные выкладки можно проделывать и с постоянной Планка и с массой частиц.

Заключение

Описанные нами переходы реальной материи в призрачную и наоборот схематичны, поскольку речь шла о частицах со спином $1/2$. Учёт частиц с иными полуцелыми спинами и бозонов требует отыскания призрачных решений соответствующих уравнений. Призрачные фотоны были найдены в статье [20], но это потребовало использования интуиционистской математики.

Является ли наша Вселенная такой, как представлено в статье, – это вопрос, для ответа на который необходимы дополнительные исследования. Мы представили лишь возможную модель, но ведь и три вселенные Фридмана – это также всего лишь модели, одни из многочисленных решений уравнений Эйнштейна.

Литература

1. Brukner S., Taylor S., Cheung S., Vedral V. Quantum Entanglement in Time. URL: <https://arxiv.org/abs/quant-ph/0402127v1> (дата обращения: 23.01.2024).
2. Nowakowski M. Quantum Entanglement in Time // AIP Conference Proceedings. 2017. Vol. 1841, Art. 020007. DOI: 10.1063/1.4982771.
3. Bartkiewicz K., Černoč A., Lemr K., Miranowicz A., Nori F. Experimental temporal quantum steering // Scientific Reports. 2016. Vol. 6. Art. 38076. URL: <https://www.nature.com/articles/srep38076> (дата обращения: 23.01.2024).
4. Гуц А.К. Заполненное пространство, логика и парадокс дедушки // Пространство, время и фундаментальные взаимодействия. 2023. Вып. 2 (42). С. 28–35.
5. Гуц А.К. О машине времени и парадоксе дедушки // Математические структуры и моделирование. 2022. № 2 (62). С. 167–172.
6. Loos H.G. Free Ghost Gauge Fields // Nuovo cimento. 1967. Vol. LII A, No. 4. P. 1085–1091.
7. Griffiths J.B. Gravitational radiation and neutrinos // Communications in Mathematical Physics. 1972. Vol. 28. P. 295–299.
8. Гуц А.К. Частицы-призраки, сцепленность исторических эпох и машина времени // Математические структуры и моделирование. 2020. № 3 (55). С. 12–21.
9. Novello M. Ghost basis for neutrino // Physics Letters A. 1976. Vol. 58, Iss. 2. P. 75–76.
10. Palesheva E.V. Ghost spinors, shadow electrons and the Deutsch Multiverse. URL: <https://arxiv.org/abs/gr-qc/0108017> (дата обращения: 23.01.2024).
11. Гуц А.К. Элементы теории времени. М.: УРСС, 2012.
12. Дойч Д. Структура реальности. Ижевск: Регулярная и хаотическая динамика, 2001.
13. Владимиров Ю.С. Метафизика. М.: БИНОМ. Лаборатория знаний, 2009.
14. Лебедев Ю.А. Неоднозначное мироздание. Кострома, 2000.
15. Гегель Г.В. Наука логики: в 3 т. М.: Мысль, 1970. Т. 1.
16. Александров А.Д. Пространство и время в современной физике // Проблемы науки и позиция ученого. Л.: Наука, 1988.

17. Палешева Е. В. Спинорные духи, теневые электроны и мультиверс Дойча // Математические структуры и моделирование. 2001. Вып. 8. С. 66–75.
18. Kock A. Synthetic Differential Geometry. Cambridge University Press, 1981.
19. Moerdijk I., Reyes G.E. Models for Smooth Infinitesimal Analysis. Springer-Verlag, 1991.
20. Guts A.K. Topos-theoretic model of the Deutsch multiverse. URL: <https://arxiv.org/abs/physics/0203071> (дата обращения: 23.01.2024).

PHYSICAL STATE OF PAST AND PRESENT OBJECTS

A.K. Guts

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

International Innovation University, Sochi, Russia

Abstract. The theory of absolute space-time, on which the theory of relativity is based, thanks to Minkowski, says that the events of the past, present and future are equal in relation to their existence, i.e. they always exist and never disappear. The article examines the question of the physical state of their existence. A formalism is being built within the framework of the Relativity theory, according to which, since we do not see or interact with objects of the past and future, their states represent the so-called ghostly matter, characterized by a zero energy-momentum tensor. With the passage of coordinate (global) time, ghostly matter passes into the present and is discovered as reality, and then again passes into ghostly and becomes an object of the past.

Keywords: theory of absolute space-time, past, future, ghostly matter.

Дата поступления в редакцию: 26.02.2024

ЗАДАЧА КОШИ ДЛЯ ДВУМЕРНЫХ ТЕЧЕНИЙ ГАЗА, ГРАВИТИРУЮЩЕГО ПО НЬЮТОНУ В ПРОСТРАНСТВЕ R^2

С.Л. Дерябин

д.ф.-м.н., профессор, e-mail: SDeryabin@usurt.ru

А.П. Садов

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

Уральский государственный университет путей сообщения, Екатеринбург, Россия

Аннотация. В работе рассматриваются изэнтропические течения идеального газа, гравитирующего по Ньютону. В качестве математических моделей получены двумерные интегро-дифференциальные системы уравнений газовой динамики для политропного газа. Для полученных уравнений поставлена задача Коши во всём пространстве R^2 . Решение задачи построено в виде степенных рядов. Коэффициенты рядов найдены при решении алгебраических уравнений с интегральными правыми частями. Получены ограничения на начальные условия задачи Коши, при которых сходятся несобственные интегралы в правых частях алгебраических уравнений.

Ключевые слова: газ, гравитирующий по Ньютону, интегро-дифференциальная система уравнений газовой динамики, задача Коши, степенные ряды, несобственные интегралы .

Введение

В работе [1] для описания течений газа, гравитирующего по Ньютону, получены системы интегро-дифференциальных уравнений. При исследовании этих уравнений в [2] были найдены трёхмерные стационарные течения самогравитирующего газа. Для одномерных течений самогравитирующего газа [2–5] удалось построить дифференциальные модели уравнений газовой динамики и решить основные начально-краевые задачи об истечении газа в вакуум. В работе [6] была сделана попытка построения решения задачи Коши для трёхмерной интегро-дифференциальной системы уравнений газовой динамики. Решение строилось в виде ряда. К сожалению, удалось выписать и проанализировать только первый член ряда. Для построения решения задачи Коши в ограниченной области для самогравитирующего газа [7] была использована дифференциальная система уравнений. Однако не было доказано, что использованная дифференциальная система уравнений эквивалентна интегро-дифференциальной системе уравнений газовой динамики.

В данной работе будет исследоваться задача Коши для двумерных интегро-дифференциальных систем уравнений газовой динамики с начальными данными, поставленными во всём пространстве R^2 .

1. Построение математической модели

Будут рассматриваться изэнтропические течения газа со следующими искомыми газо-динамическими параметрами: u, v, w – декартовы координаты вектора скорости газа; ρ – плотность газа.

Система уравнений, описывающая изэнтропические течения газа, гравитирующего по Ньютону, имеет вид [1]

$$\begin{aligned} \rho_t + u\rho_x + v\rho_y + w\rho_z + \rho(u_x + v_y + w_z) &= 0, \\ u_t + uu_x + vu_y + wu_z + \frac{1}{\rho}p_x &= F_1, \\ v_t + uv_x + vv_y + wv_z + \frac{1}{\rho}p_y &= F_2, \\ w_t + uw_x + vw_y + ww_z + \frac{1}{\rho}p_z &= F_3. \end{aligned} \tag{1}$$

Здесь

$$\begin{aligned} F_1 &= G \iiint_{\Omega} \frac{\rho(t, \xi_1, \xi_2, \varsigma)(x - \xi_1)}{(\sqrt{(x - \xi_1)^2 + (y - \xi_2)^2 + (z - \varsigma)^2})^3} d\xi_1 d\xi_2 d\varsigma; \\ F_2 &= G \iiint_{\Omega} \frac{\rho(t, \xi_1, \xi_2, \varsigma)(y - \xi_2)}{(\sqrt{(x - \xi_1)^2 + (y - \xi_2)^2 + (z - \varsigma)^2})^3} d\xi_1 d\xi_2 d\varsigma; \\ F_3 &= G \iiint_{\Omega} \frac{\rho(t, \xi_1, \xi_2, \varsigma)(z - \varsigma)}{(\sqrt{(x - \xi_1)^2 + (y - \xi_2)^2 + (z - \varsigma)^2})^3} d\xi_1 d\xi_2 d\varsigma, \end{aligned}$$

где p – давление; $\Omega = R^3$ – область, занимаемая газом; $G = 6,67 \cdot 10^{-11} \frac{\text{Нм}^2}{\text{кг}^2}$ – гравитационная постоянная.

Для построения двумерных течений газа предположим, что область Ω является бесконечным цилиндром, в основании которого плоскость $D = R^2$.

Пусть во всех точках области Ω заданы следующие параметры газа:

$$\rho = \rho(t, x, y), \quad u = u(t, x, y), \quad v = v(t, x, y), \quad w = 0.$$

Тогда гравитирующая сила $\vec{F} = \{F_1, F_2, F_3\}$ вычисляется по формулам

$$\begin{aligned}
F_1 &= G \iint_D \rho(t, \xi_1, \xi_2)(x - \xi_1) d\xi_1 d\xi_2 \int_{-\infty}^{+\infty} \frac{1}{(\sqrt{(x - \xi_1)^2 + (y - \xi_2)^2 + (z - \varsigma)^2})^3} d\varsigma; \\
F_2 &= G \iint_D \rho(t, \xi_1, \xi_2)(y - \xi_2) d\xi_1 d\xi_2 \int_{-\infty}^{+\infty} \frac{1}{(\sqrt{(x - \xi_1)^2 + (y - \xi_2)^2 + (z - \varsigma)^2})^3} d\varsigma; \\
F_3 &= G \iint_D \rho(t, \xi_1, \xi_2)(z - \varsigma) d\xi_1 d\xi_2 \int_{-\infty}^{+\infty} \frac{1}{(\sqrt{(x - \xi_1)^2 + (y - \xi_2)^2 + (z - \varsigma)^2})^3} d\varsigma.
\end{aligned} \tag{2}$$

В первых двух равенствах системы (2) сделаем замену переменных

$$z - \varsigma = \sqrt{(x - \xi_1)^2 + (y - \xi_2)^2} \operatorname{tg} u.$$

Получим

$$\begin{aligned}
F_1 &= -G \iint_D \rho(t, \xi_1, \xi_2)(x - \xi_1) d\xi_1 d\xi_2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\cos u}{(x - \xi_1)^2 + (y - \xi_2)^2} du; \\
F_2 &= -G \iint_D \rho(t, \xi_1, \xi_2)(y - \xi_2) d\xi_1 d\xi_2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\cos u}{(x - \xi_1)^2 + (y - \xi_2)^2} du.
\end{aligned}$$

После интегрирования имеем

$$\begin{aligned}
F_1 &= -2G \iint_D \frac{\rho(t, \xi_1, \xi_2)(x - \xi_1)}{(x - \xi_1)^2 + (y - \xi_2)^2} d\xi_1 d\xi_2; \\
F_2 &= -2G \iint_D \frac{\rho(t, \xi_1, \xi_2)(y - \xi_2)}{(x - \xi_1)^2 + (y - \xi_2)^2} d\xi_1 d\xi_2.
\end{aligned}$$

Вычисляя последний интеграл в F_3 , получим

$$F_3 = G \iint_D \rho(t, \xi_1, \xi_2) \frac{1}{\sqrt{(x - \xi_1)^2 + (y - \xi_2)^2 + (z - \varsigma)^2}} \Big|_{-\infty}^{+\infty} d\xi_1 d\xi_2 = 0.$$

Таким образом, четвёртое уравнение системы (1) выполняется тождественно, и система (1) переписывается в виде

$$\begin{aligned} \rho_t + u\rho_x + v\rho_y + \rho(u_x + v_y) &= 0, \\ u_t + uu_x + vv_y + \frac{1}{\rho}p_x &= -2G \iint_D \frac{\rho(t, \xi_1, \xi_2)(x - \xi_1)}{(x - \xi_1)^2 + (y - \xi_2)^2} d\xi_1 d\xi_2; \\ v_t + uv_x + vv_y + \frac{1}{\rho}p_y &= -2G \iint_D \frac{\rho(t, \xi_1, \xi_2)(y - \xi_2)}{(x - \xi_1)^2 + (y - \xi_2)^2} d\xi_1 d\xi_2. \end{aligned} \quad (3)$$

Не нарушая общности, уравнение состояния политропного газа возьмём в виде $p = \frac{\rho^\gamma}{\gamma}$, $\gamma = \text{const} > 1$. Тогда система (3) для изэнтропических течений газа будет иметь вид

$$\begin{aligned} \rho_t + u\rho_x + v\rho_y + \rho(u_x + v_y) &= 0, \\ u_t + uu_x + vv_y + \rho^{\gamma-2}\rho_x &= \\ &= -2G \iint_D \frac{\rho(t, \xi_1, \xi_2)(x - \xi_1)}{(x - \xi_1)^2 + (y - \xi_2)^2} d\xi_1 d\xi_2; \\ v_t + uv_x + vv_y + \rho^{\gamma-2}\rho_y &= \\ &= -2G \iint_D \frac{\rho(t, \xi_1, \xi_2)(y - \xi_2)}{(x - \xi_1)^2 + (y - \xi_2)^2} d\xi_1 d\xi_2. \end{aligned} \quad (4)$$

Замечание 1. Заметим, что система (4) для произвольного числа γ не является аналитической. Полагая, что $(\gamma - 2) = k$ – целое положительное число, получим счётный набор $\gamma = (2+k) = \{2; 3; 4; 5; 6; 7; \dots\}$, для которых функция $\rho^{\gamma-2}$ является аналитической. Заметим, что для воды $\gamma = 7$, т. е. $k = 5$. При этом для воздуха $\gamma = 1,4$ в этот набор не входит.

Если за неизвестную функцию взять c – скорость звука газа $\left(c^2 = \frac{dp}{d\rho} = \rho^{\gamma-1}\right)$, то система (4) будет иметь вид:

$$\begin{aligned}
c_t + uc_x + vc_y + \frac{\gamma - 1}{2}c(u_x + v_y) &= 0, \\
u_t + uu_x + vv_y + \frac{2}{\gamma - 1}cc_x &= \\
= -2G \iint_D c^{\gamma - 1}(t, \xi_1, \xi_2) \frac{x - \xi_1}{(x - \xi_1)^2 + (y - \xi_2)^2} d\xi_1 d\xi_2; \\
v_t + uv_x + vv_y + \frac{2}{\gamma - 1}cc_y &= \\
= -2G \iint_D c^{\gamma - 1}(t, \xi_1, \xi_2) \frac{y - \xi_2}{(x - \xi_1)^2 + (y - \xi_2)^2} d\xi_1 d\xi_2.
\end{aligned} \tag{5}$$

Интегралы в правой части системы (5) запишем в полярной системе координат:

$$\xi_1 - x = r \cos \varphi, \quad \xi_2 - y = r \sin \varphi, \quad d\xi_1 d\xi_2 = r dr d\varphi.$$

Соответственно, пределы интегрирования будут иметь вид

$$D: 0 \leq \varphi \leq 2\pi, \quad 0 \leq r < \infty.$$

Тогда система (5) перепишется в виде

$$\begin{aligned}
c_t + uc_x + vc_y + \frac{\gamma - 1}{2}c(u_x + v_y) &= 0, \\
u_t + uu_x + vv_y + \frac{2}{\gamma - 1}cc_x &= \\
= -2G \int_0^{2\pi} \left[\int_0^{\infty} c^{\gamma - 1}(t, x + r \cos \varphi, y + r \sin \varphi) dr \right] \cos \varphi d\varphi; \\
v_t + uv_x + vv_y + \frac{2}{\gamma - 1}cc_y &= \\
= -2G \int_0^{2\pi} \left[\int_0^{\infty} c^{\gamma - 1}(t, x + r \cos \varphi, y + r \sin \varphi) dr \right] \sin \varphi d\varphi.
\end{aligned} \tag{6}$$

Замечание 2. В результате такой замены мы получили аналитическую систему уравнений газовой динамики, но для произвольного числа γ подынтегральная функция $c^{\frac{2}{\gamma-1}}$ не является аналитической. Предполагая, что $\frac{2}{\gamma-1} = n$ – натуральное число, получим счётный набор $\gamma = \left(1 + \frac{2}{n}\right) = \left[3; 2; \frac{5}{3}; \frac{3}{2}; 1, 4; \dots\right]$, для которых подынтегральная функция является аналитической. Заметим, что для воздуха $\gamma = 1, 4$, т. е. $n = 5$.

2. Построение решения задачи Коши

Система (6) с учётом замечания 2 будет иметь вид

$$c_t + uc_x + vc_y + \frac{1}{n}c(u_x + v_y) = 0,$$

$$u_t + uu_x + vv_y + ncc_x =$$

$$= -2G \int_0^{2\pi} \left[\int_0^\infty c^n(t, x + r \cos \varphi, y + r \sin \varphi) \right] \cos \varphi d\varphi; \tag{7}$$

$$v_t + uv_x + vv_y + ncc_y =$$

$$= -2G \int_0^{2\pi} \left[\int_0^\infty c^n(t, x + r \cos \varphi, y + r \sin \varphi) \right] \sin \varphi d\varphi.$$

Пусть при $t = t_0$ заданы начальные условия

$$c(t_0, x, y) = c_0(x, y), \quad u(t_0, x, y) = u_0(x, y), \quad v(t_0, x, y) = v_0(x, y). \tag{8}$$

Далее будет предполагаться, что функции $c_0(x, y)$, $u_0(x, y)$, $v_0(x, y)$ ограничены во всей области R^2 и интеграл

$$\int_0^\infty c_0^n(x + r \cos \varphi, y + r \sin \varphi) dr \tag{9}$$

сходится.

Построим решение задачи (7), (8) в виде ряда по степеням t

$$\mathbf{f}(t, x, y) = \sum_{k=0}^\infty \mathbf{f}_k(x, y) \frac{(t - t_0)^k}{k!}, \quad \mathbf{f} = \{c, u, v\}. \tag{10}$$

Нулевые коэффициенты ряда (10) находятся из начальных условий (8).

В системе (7) положим $t = t_0$ и получим первые коэффициенты ряда (10):

$$\begin{aligned}
 c_1 &= -u_0 c_{0x} - v_0 c_{0y} - \frac{1}{n} c_0 (u_{0x} + v_{0y}); \\
 u_1 &= -u_0 u_{0x} - v_0 u_{0y} - n c_0 c_{0x} - \\
 &\quad - 2G \int_0^{2\pi} \left[\int_0^{\infty} c_0^n (x + r \cos \varphi, y + r \sin \varphi) \right] \cos \varphi d\varphi; \\
 v_1 &= -u_0 v_{0x} - v_0 v_{0y} - n c_0 c_{0y} - \\
 &\quad - 2G \int_0^{2\pi} \left[\int_0^{\infty} c_0^n (x + r \cos \varphi, y + r \sin \varphi) \right] \sin \varphi d\varphi.
 \end{aligned} \tag{11}$$

Продифференцируем систему (7) по t , положим $t = t_0$ и получим вторые коэффициенты ряда (10)

$$\begin{aligned}
 c_2 &= -u_1 c_{0x} - u_0 c_{1x} - v_0 c_{0y} - v_1 c_{1y} - \frac{\gamma - 1}{2} c_1 (u_{0x} + v_{0y}) - \\
 &\quad - \frac{1}{n} c_0 (u_{1x} + v_{1y}); \\
 u_2 &= -u_1 u_{0x} - u_0 u_{1x} - v_1 u_{0y} - v_0 u_{1y} - n (c_1 c_{0x} + c_0 c_{1x}) - \\
 &\quad - 2G \int_0^{2\pi} \left[\int_0^{\infty} n c_0^{n-1} c_1 dr \right] \cos \varphi d\varphi; \\
 v_2 &= -u_1 v_{0x} - u_0 v_{1x} - v_1 v_{0y} - v_0 v_{1y} - n c_0 c_{0y} - \\
 &\quad - 2G \int_0^{2\pi} \left[\int_0^{\infty} n c_0^{n-1} c_1 dr \right] \sin \varphi d\varphi.
 \end{aligned}$$

Далее будет предполагаться, что интеграл

$$\int_0^{\infty} n c_0^{n-1} (x + r \cos \varphi, y + r \sin \varphi) c_1 (x + r \cos \varphi, y + r \sin \varphi) dr \tag{12}$$

сходится.

Остальные коэффициенты ряда получаются рекуррентным образом с помощью дифференцирования системы (7) по t подстановки в полученные выражения $t = t_0$ и ранее вычисленных коэффициентов ряда.

Продифференцируем систему (7) k раз по t , положим $t = t_0$ и получим

$$\begin{aligned}
 c_{k+1} &= F_{1k}(x, y); \\
 u_{k+1} &= F_{2k}(x, y) - 2G \int_0^{2\pi} \left[\int_0^{\infty} P_k(x, y, r, \varphi) \right] \cos \varphi d\varphi; \\
 v_{k+1} &= F_{3k}(x, y) - 2G \int_0^{2\pi} \left[\int_0^{\infty} P_k(x, y, r, \varphi) dr \right] \sin \varphi d\varphi.
 \end{aligned} \tag{13}$$

Здесь

$$\begin{aligned}
 F_{1k}(x, y) &= - \sum_{p=0}^k C_k^p u_p \cdot c_{x(k-p)} - \\
 &- \sum_{p=0}^k C_k^p v_p \cdot c_{y(k-p)} - \frac{\gamma - 1}{2} \sum_{p=0}^k C_k^p c_p [u_{x(k-p)} + v_{y(k-p)}]; \\
 F_{2k}(x, y) &= - \sum_{p=0}^k C_k^p u_p \cdot u_{x(k-p)} - \sum_{p=0}^k C_k^p v_p \cdot u_{y(k-p)} - \\
 &- \frac{2}{\gamma - 1} \sum_{p=0}^k C_k^p c_p c_{x(k-p)}; \\
 F_{3k}(x, y) &= - \sum_{p=0}^k C_k^p u_p \cdot v_{x(k-p)} - \sum_{p=0}^k C_k^p v_p \cdot v_{y(k-p)} - \\
 &- \frac{2}{\gamma - 1} \sum_{p=0}^k C_k^p c_p c_{y(k-p)}; \\
 P_k(x, y, r, \varphi) &= \frac{d^k}{dt^k} [c^n(t, x + r \cos \varphi, y + r \sin \varphi)]_{t=t_0}.
 \end{aligned} \tag{14}$$

Далее будет предполагаться, что интегралы

$$\int_0^{\infty} P_k(x, y, r, \varphi) dr \tag{15}$$

сходятся.

При сделанных предположениях рекуррентные соотношения для построения решения задачи Коши получены.

Лемма. Пусть в окрестности точки $M_0(x_0, y_0)$ заданы функции $u_0(x, y) = \frac{u_{00}(x, y)}{x \cdot y}$, $v_0(x, y) = \frac{v_{00}(x, y)}{x \cdot y}$, $c_0(x, y) = \frac{c_{00}(x, y)}{x \cdot y}$. Если функции $c_{00}(x, y)$, $u_{00}(x, y)$, $v_{00}(x, y)$ и их частные производные любого порядка ограничены в области R^2 , тогда несобственные интегралы (15) сходятся.

Доказательство. Лемма доказывается индукцией по k . База индукции:

$$\int_0^{\infty} c_0^n(x + r \cos \varphi, y + r \sin \varphi) dr = \lim_{R \rightarrow \infty} \int_0^R \frac{c_{00}^n(x + r \cos \varphi, y + r \sin \varphi)}{(x + r \cos \varphi)^n \cdot (y + r \sin \varphi)^n} dr.$$

Заметим, что n – фиксированное натуральное число. Следовательно, функция $c_{00}^n(x + r \cos \varphi, y + r \sin \varphi)$ ограничена в области R^2 , т. е.

$$|c_{00}^n(x + r \cos \varphi, y + r \sin \varphi)| < A.$$

Тогда

$$\begin{aligned} & \int_0^R \frac{c_{00}^n(x + r \cos \varphi, y + r \sin \varphi)}{(x + r \cos \varphi)^n \cdot (y + r \sin \varphi)^n} dr < \\ & < \int_0^R \frac{A}{(x + r \cos \varphi)^n \cdot (y + r \sin \varphi)^n} dr = J_n. \end{aligned}$$

Вычислим интеграл J_n .

При $n = 1$

$$J_1 = A \int_0^R \frac{1}{(x + r \cos \varphi) \cdot (y + r \sin \varphi)} dr =$$

$$= \frac{A}{y \cos \varphi - x \sin \varphi} \ln \left| \frac{(x + R \cos \varphi)y}{(y + R \sin \varphi)x} \right|;$$

$$\lim_{R \rightarrow \infty} J_1 = \lim_{R \rightarrow \infty} \frac{A}{y \cos \varphi - x \sin \varphi} \ln \left| \frac{(x + R \cos \varphi)y}{(y + R \sin \varphi)x} \right| =$$

$$= \frac{A}{y \cos \varphi - x \sin \varphi} \ln \left| \frac{y \cos \varphi}{x \sin \varphi} \right|$$

интеграл (9) сходится.

При $n > 1$ порядок J_n будет $\frac{1}{r^{2n}}$, и это гарантирует сходимость интеграла (9).

Дальнейшее доказательство проведём для $n = 1$.

Рассмотрим интегралы

$$I_1 = \int_0^{2\pi} \frac{A \cos \varphi}{y \cos \varphi - x \sin \varphi} \ln \left| \frac{y \cos \varphi}{x \sin \varphi} \right| d\varphi;$$

$$I_2 = \int_0^{2\pi} \frac{A \sin \varphi}{y \cos \varphi - x \sin \varphi} \ln \left| \frac{y \cos \varphi}{x \sin \varphi} \right| d\varphi.$$

Перепишем их в виде

$$I_1 = \frac{A}{y} \int_0^{2\pi} \frac{\operatorname{ctg} \varphi}{\operatorname{ctg} \varphi - \frac{x}{y}} \ln \left| \frac{y}{x} \operatorname{ctg} \varphi \right| d\varphi, \quad I_2 = \frac{A}{x} \int_0^{2\pi} \frac{\operatorname{tg} \varphi}{\operatorname{tg} \varphi - \frac{y}{x}} \ln \left| \frac{x}{y} \operatorname{tg} \varphi \right| d\varphi.$$

Это несобственные интегралы от функций, разрывных в точках: $\varphi = 0$; $\varphi = \frac{\pi}{2}$; $\varphi = \pi$; $\varphi = \frac{3\pi}{2}$ $\varphi = \operatorname{arctg} \frac{y}{x}$.

Далее интегралы I_1, I_2 рассмотрим в пределах от 0 до $\frac{\pi}{2}$.

В интеграле I_1 сделаем замену переменных $z = \operatorname{ctg} \varphi$, в интеграле I_2 – замена $z = \operatorname{tg} \varphi$. Получим

$$I_1 = \frac{A}{y} \int_0^{\infty} \frac{z}{\left(z - \frac{x}{y}\right)(1 + z^2)} \ln \left(\frac{y}{x} z\right) dz;$$

$$I_2 = \frac{A}{x} \int_0^{\infty} \frac{z}{\left(z - \frac{y}{x}\right)(1 + z^2)} \ln \left(\frac{x}{y} z\right) dz.$$

Исследуем интеграл I_1 в точке $z = \frac{x}{y}$. Интеграл рассмотрим в пределах $\left[\frac{x}{y} - a, \frac{x}{y} + b\right]$, $a \in \left(0, \frac{x}{y}\right)$, $b \in \left(\frac{x}{y}, \infty\right)$.

$$I_1 = \frac{A}{y} \int_{\frac{x}{y}-a}^{\frac{x}{y}} \frac{z}{\left(z - \frac{x}{y}\right)(1 + z^2)} \ln \left(\frac{y}{x} z\right) dz +$$

$$+ \frac{A}{y} \int_{\frac{x}{y}}^{\frac{x}{y}+b} \frac{z}{\left(z - \frac{x}{y}\right)(1 + z^2)} \ln \left(\frac{y}{x} z\right) dz.$$

Сделаем замену переменных $U = \frac{y}{x}z$, $z = \frac{x}{y}U$. Будем иметь

$$I_1 = \frac{Ax}{y^2} \int_{1-\frac{a}{x}}^1 \frac{U}{(U-1) \left(1 + \left(\frac{x}{y}U\right)^2\right)} \ln U dU +$$

$$+ \frac{Ax}{y^2} \int_1^{1+\frac{b}{x}} \frac{U}{(U-1) \left(1 + \left(\frac{x}{y}U\right)^2\right)} \ln U dU.$$

Или, если $V = U - 1$,

$$I_1 = \frac{Ax}{y^2} \int_{-\frac{a}{x}}^0 \frac{V+1}{V \left(1 + \frac{x^2}{y^2}(V+1)^2\right)} \ln(1+V) dV +$$

$$+ \frac{Ax}{y^2} \int_0^{\frac{b}{x}} \frac{V+1}{V \left(1 + \frac{x^2}{y^2}(V+1)^2\right)} \ln(1+V) dV.$$

Поскольку $\ln(1+V) \leq V$, то справедлива оценка

$$I_1 \leq \frac{Ax}{y^2} \int_{-\frac{a}{x}}^0 \frac{V+1}{1 + \frac{x^2}{y^2}(V+1)^2} dV + \frac{A}{y} \int_0^{\frac{b}{x}} \frac{V+1}{1 + \frac{x^2}{y^2}(V+1)^2} dV$$

или

$$I_1 \leq \frac{A}{2x} \int_{-\frac{a}{x}}^0 \frac{d \left(1 + \frac{x^2}{y^2}(V+1)^2\right)}{1 + \frac{x^2}{y^2}(V+1)^2} + \frac{A}{2x} \int_0^{\frac{b}{x}} \frac{d \left(1 + \frac{x^2}{y^2}(V+1)^2\right)}{1 + \frac{x^2}{y^2}(V+1)^2}.$$

Интегрируя, имеем

$$I_1 \leq \frac{A}{2x} \left[\ln \left(1 + \frac{x^2}{y^2}(V+1)^2\right) \right] \Big|_{-\frac{a}{x}}^0 + \ln \left(1 + \frac{x^2}{y^2}(V+1)^2\right) \Big|_0^{\frac{b}{x}}$$

или

$$I_1 \leq \frac{A}{2x} \ln \frac{1 + \frac{x^2}{y^2} \left(1 + b \frac{y}{x}\right)^2}{1 + \frac{x^2}{y^2} \left(1 - a \frac{y}{x}\right)^2}.$$

Аналогично получаем оценку для I_2 :

$$I_2 \leq \frac{A}{2y} \ln \frac{1 + \frac{y^2}{x^2} \left(1 + b \frac{x}{y}\right)^2}{1 + \frac{y^2}{x^2} \left(1 - a \frac{x}{y}\right)^2}.$$

Следовательно, интеграл I_1 в точке $z = \frac{x}{y}$ и интеграл I_2 в точке $z = \frac{y}{x}$ сходятся.

Исследуем интеграл I_1 в точке $z = 0$. Рассмотрим его в пределах $[0, a]$ и после замены $U = \frac{y}{x}z$, $z = \frac{x}{y}U$ получим

$$I_1 = \frac{Ax}{y^2} \int_0^{\frac{ay}{x}} \frac{U}{(U-1) \left(1 + \left(\frac{xU}{y}\right)^2\right)} \ln U dU.$$

В окрестности точки $U = 0$ справедливо неравенство $\ln U \leq \frac{1}{U}$. Тогда

$$I_1 \leq \frac{Ax}{y^2} \int_0^{\frac{ay}{x}} \frac{U^2}{(U-1) \left(1 + \left(\frac{xU}{y}\right)^2\right)} dU \leq \frac{Ax}{y^2} \int_0^{\frac{ay}{x}} \frac{U}{1 + \left(\frac{xU}{y}\right)^2} dU.$$

Интегрируя, имеем

$$I_1 \leq \frac{A}{2x} \ln \left(1 + \frac{x^2}{y^2} U^2\right) \Big|_0^{\frac{ay}{x}} = \frac{A}{2x} \ln 2.$$

Оценка для I_2 будет иметь вид

$$I_2 \leq \frac{A}{2y} \ln 2.$$

Следовательно, интегралы I_1 и I_2 в точке $z = 0$ сходятся.

Исследуем интеграл I_1 при $z \rightarrow \infty$. Рассмотрим его в пределах $[b, \infty]$ и после замены $U = \frac{y}{x}z$, $z = \frac{x}{y}U$ получим

$$I_1 = \frac{Ax}{y^2} \int_{\frac{by}{x}}^{\infty} \frac{U}{(U-1) \left(1 + \left(\frac{xU}{y}\right)^2\right)} \ln U dU.$$

При $U \rightarrow \infty$ справедливо неравенство $\ln U \leq \sqrt{U}$. Тогда

$$I_1 \leq \frac{Ax}{y^2} \int_{b\frac{y}{x}}^{\infty} \frac{U\sqrt{U}}{(U-1)\left(1 + \left(\frac{x}{y}U\right)^2\right)} dU \leq \frac{A}{x} \int_{b\frac{y}{x}}^{\infty} \frac{1}{(U-1)\sqrt{U}} dU.$$

Делая замену переменных $U = w^2$, получим

$$I_1 \leq \frac{2A}{x} \int_{\sqrt{b\frac{y}{x}}}^{\infty} \frac{1}{w^2-1} dw = \frac{A}{x} \ln \left| \frac{w-1}{w+1} \right| \Big|_{\sqrt{b\frac{y}{x}}}^{\infty} = \frac{A}{x} \ln \left| \frac{b\frac{y}{x} + 1}{b\frac{y}{x} - 1} \right|.$$

Оценка для I_2 будет иметь вид

$$I_2 \leq \frac{A}{y} \ln \left| \frac{b\frac{x}{y} + 1}{b\frac{x}{y} - 1} \right|.$$

При $U \rightarrow \infty$ интегралы I_1 и I_2 сходятся.

База индукции доказана.

Заметим, что условия леммы и формулы (11), (12) гарантируют, что c_1, c_2 имеет такой же порядок по x, y , как и c_0 .

Делаем индуктивное предположение, что, при $l < k$, c_l имеет такой порядок по x, y , как и c_0 . Тогда из формул (13) и условий леммы получаем, что интегралы (15) сходятся. Лемма доказана. ■

Заключение

1. В работе построена математическая модель для описания двумерных течений газа, гравитирующего по Ньютону.
2. Для интегро-дифференциальной системы уравнений газовой динамики поставлена задача Коши, решение которой построено в виде степенного ряда.
3. Коэффициенты ряда получены с помощью рекуррентных соотношений из решения алгебраически уравнений с интегральными правыми частями.
4. Получены ограничения на начальные условия задачи Коши, при которых сходятся несобственные интегралы в правых частях алгебраических уравнений.

Таким образом, выполнено аналитическое исследование для дальнейшего численного моделирования гравитационных волн на большой промежуток времени.

Литература

1. Ламб Г. Гидродинамика. М.; Л.: ОГИЗ, 1947.
2. Дерябин С.Л., Чуев Н.П. Сферически симметричное истечение самогравитирующего идеального газа в вакуум // Прикладная математика и механика. 1994. Т. 58, вып. 2. С. 77–84.
3. Баутин С.П., Дерябин С.Л. Математическое моделирование истечения идеального газа в вакуум. Новосибирск: Наука, 2005.
4. Дерябин С.Л. Одномерное истечение самогравитирующего идеального газа в вакуум // Вычислительные технологии. 2003. Т. 8, № 4. С. 32–44.
5. Дерябин С.Л., Садов А.П. Математическое моделирование течений самогравитирующего газа с помощью стационарных автомодельных переменных // Вестник Уральского государственного университета путей сообщения. 2022. № 3 (55). С. 15–22.
6. Чуев Н.П. Аналитический метод исследования пространственных задач динамики самогравитирующего газа // Вычислительные технологии. 1998. Т. 3, № 1. С. 79–89.
7. Дерябин С.Л., Чуев Н.П. Построение трёхмерных течений самогравитирующего идеального газа, непрерывно примыкающих к вакууму // Вестник Уральского государственного университета путей сообщения. 2012. № 2 (14). С. 4–13.

THE CAUCHY PROBLEM FOR TWO-DIMENSIONAL GAS FLOWS, GRAVITATING BY NEWTON IN THE SPACE OF R^2 **S.L. Deryabin**

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

A.P. Sadov

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

Ural State University of Railway Engineering, Yekaterinburg, Russia

Abstract. The paper considers isentropic flows of an ideal gas gravitating by Newton. Two-dimensional integro-differential systems of equations of gas dynamics for a polytropic gas are obtained as mathematical models. For the obtained equations, the Cauchy problem is posed in the entire space R^2 . The solution of the problem is constructed in the form of power series. The coefficients of the series are found in solving algebraic equations with integral right-hand sides. Restrictions are obtained on the initial conditions of the Cauchy problem, under which improper integrals converge in the right-hand sides of algebraic equations.

Keywords: gas engraving according to Newton, integro-differential system of equations of gas dynamics, Cauchy problem, power series, improper integrals.

Дата поступления в редакцию: 24.12.2023

ПРИМЕНЕНИЕ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ДИАГНОСТИКИ ФЕНОТИПОВ ЗАБОЛЕВАНИЯ ЖЕЛУДОЧНО-КИШЕЧНОГО ТРАКТА

А.А. Березин¹

аспирант, e-mail: andreyberezin55@gmail.com

С.А. Агалаков²

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

¹Омский государственный технический университет, Омск, Россия

²Омский государственный университет им. Ф.М. Достоевского, Омск, Россия

Аннотация. Работа посвящена задаче диагностики фенотипов заболеваний желудочно-кишечного тракта с помощью моделей машинного обучения. Главной целью является поиск модели, решающей задачу классификации пациентов медицинского учреждения. В ходе работы сначала были рассмотрены простые модели, такие как логистическая регрессия и кластерный анализ. Ввиду неэффективности этих методов был выполнен поиск более сложных моделей и методов анализа данных: деревья решений, ансамбли деревьев, искусственные нейронные сети. Также была проведена работа по предварительной обработке набора данных и выбору значимых признаков для классификации. В результате была найдена наилучшая модель для рассматриваемого набора данных – CatBoost, которая диагностирует фенотипы с точностью 92,85 % на тренировочной выборке и на тестовой выборке – 79,31 %. Кроме того, в работе оценивается применимость методов машинного обучения в клинической практике и определяются направления будущих исследований для более точной диагностики.

Ключевые слова: машинное обучение, задача классификации, логистическая регрессия, нейронные сети, ансамбли решающих деревьев, бустинг.

Введение

В настоящее время машинное обучение и анализ данных становятся всё более актуальными в различных областях: в экономике, науке, здравоохранении, технике, маркетинге и т. д. Модели машинного обучения могут решать многие проблемы [1], но в эпоху пандемии и увеличения числа видов заболеваний как никогда актуальны задачи диагностики в здравоохранении. В общем и целом, медицина – это одна из самых важных сфер жизни людей, поэтому в данной работе рассматривается задача классификации фенотипов заболеваний желудочно-кишечного тракта (далее – ЖКТ) с помощью моделей машинного обучения.

Данные для работы были предоставлены Омским государственным медицинским университетом, они содержат информацию о реальных пациентах. Этот факт при-

даёт особую ценность и значимость проведённому исследованию, обогащая его реальным клиническим контекстом. В результате данного анализа предпринимается попытка улучшить процесс диагностики, что может иметь непосредственное влияние на практическую медицинскую практику и благополучие пациентов.

В качестве инструментов для практической работы с методами машинного обучения были использованы: язык Python, компоненты библиотек scikit-learn, Keras, NumPy.

1. Описание задачи

В данной работе поставлена определённая задача классификации фенотипов заболевания ЖКТ. Задача состоит в следующем. Есть результаты анкетирования пациентов и их медицинские показатели, представленные в таблице (112 признаков, 281 пациент), и на основе этих данных нужно отнести каждого пациента к одному из шести фенотипов заболевания ЖКТ: здоровые лица, ПИ-СРК, с ожирением, коморбидность, эссенциальный и смешанный фенотип.

Все исходные признаки разделены на пять групп вопросов/показателей: пищевые привычки (P), уровень стресса (S), симптомы болезни (B), показатели качества жизни (K), лабораторные маркеры (M).

Данная задача является примером задачи машинного обучения «с учителем», т. е. на вход модели будут поступать данные с правильными ответами, что необходимо для проверки качества обучения модели классификации. В качестве основной метрики будет использоваться *точность*.

Точность – это доля правильно классифицированных объектов среди всех объектов.

1.1. Описание данных

Группа «Пищевые привычки» насчитывает 70 признаков. Это очень большая группа вопросов, где людей просили ответить на вопросы относительно их предпочтений в еде и о приёме пищи. Например:

- Вы принимаете пищу примерно в одно и то же время?
- В течение последнего года пытались ли Вы изменить свои привычки питания в сторону здорового образа жизни – употреблять меньше жира?
- Сколько чашек кофе Вы обычно выпиваете за день (шт.)?
- Довольны ли Вы количеством пищи, потребляемым за сутки?
- Сколько граммов фруктов Вы употребляете в сутки?

Группа показателей «Уровень стресса» состоит из 6 признаков. Данные показатели оценивают уровень тревоги и депрессии пациента.

Группа «Симптомы болезни» насчитывает 20 показателей, среди которых: длительность заболевания, тошнота, неприятный вкус, урчание по ходу кишечника, метеоризм, индекс массы тела, частота стула, абдоминальная боль и др.

Группа «Показатели качества жизни» насчитывает 11 показателей: физическое функционирование, жизненная активность, социальное функционирование, психическое здоровье, ролевое функционирование, обусловленное эмоциональным состоянием, и др.

Группа «Лабораторные маркеры» состоит из 5 показателей: кортизол утром и вечером, серотонин, дофамин и зонулин.

1.2. Предварительная обработка данных

В данной работе на этапе разделения выборки на тестовую и тренировочную проводится нормализация данных с помощью StandardScaler из библиотеки scikit-learn для масштабирования данных. Это распространённый подход для предобработки данных перед применением моделей машинного обучения.

StandardScaler выполняет масштабирование данных путём преобразования их в стандартное нормальное распределение. А именно, данный метод применяет следующую формулу для каждого признака:

$$z = \frac{x - u}{s};$$

где z – новое масштабированное значение признака; x – исходное значение признака; u – среднее значение признака в наборе данных; s – стандартное отклонение признака в наборе данных.

Данный способ является стандартным и, как правило, необходимым для достижения большей точности при выполнении классификации.

Также были проведены эксперименты над различными пропорциями в разделении наблюдений на тренировочную и тестовую выборки: 70 % на 30 %, 80 % на 20 %, и лучшим для данной задачи оказалось разделение 90 % на 10 %.

На этапе обучения используется кросс-валидация по 5 фолдам: тренировочная выборка разделена на 5 частей и по очереди одна из них будет использована для тестирования в процессе обучения модели. Это улучшает точность на тестовой выборке.

2. Ход работы

В самом начале работы исследование началось с применения таких простых моделей, как логистическая регрессия и кластеризация методом k -средних на разных группах признаков.

2.1. Факторный анализ

Ввиду малого количества наблюдений (281) относительно большого количества переменных в данных (112), была выдвинута идея факторного анализа. Данная идея заключается в том, что с помощью факторного анализа можно сократить размерность данных, заменяя исходные переменные новыми факторами, которые объясняют основную часть изменчивости в данных, выполняя факторный анализ по группам вопросов, а не по всем показателям сразу.

В данной работе факторный анализ проводился с помощью пакета-программы SPSS по группам признаков отдельно, т. е. отдельно в группах «пищевые привычки», «уровень стресса» и остальных. При сокращении размерности использовался метод главных компонент [2]. Для определения количества факторов использовался основной критерий: доля объяснённой дисперсии должна быть около 80 %, при этом сначала нужно было выбрать те факторы, у которых собственные значения больше единицы.

В результате факторного анализа суммарное количество переменных сократилось почти в пять раз: со 112 признаков до 24. А именно для групп «Пищевые привычки», «Уровень стресса», «Симптомы болезни», «Показатели качества жизни», «Лабораторные маркеры» удалось провести факторный анализ и значительно сократить размерность: с 70, 6, 20, 11, 5 признаков до 15, 2, 2, 3, 2 соответственно.

После того как был проведён факторный анализ и получены новые переменные, была выполнена логистическая регрессия для этих значений (рис. 1). Также для сравнения результатов логистическая регрессия была выполнена и на всех исходных переменных (рис. 2).

```
Train_size = 0.9
Выборка - X_FACTORS
Best params: {'C': 0.1, 'max_iter': 100, 'penalty': 'l2', 'solver': 'newton-cg'}
Train score: 0.5595238095238095
Test score: 0.4482758620689655
```

Рис. 1. Результаты логистической регрессии для всех новых факторов

```
Train_size = 0.9
Выборка - X_ALL
Best params: {'C': 0.1, 'max_iter': 100, 'penalty': 'l1', 'solver': 'saga'}
Train score: 0.6428571428571429
Test score: 0.5172413793103449
```

Рис. 2. Результаты логистической регрессии для исходных признаков

Результаты логистической регрессии на новых факторах и на всём наборе данных оказались неудовлетворительными: 44,83 % точности для тестового набора на новых факторных переменных, 51,72 % – на всех переменных. Получается, что результаты на новых факторах ещё хуже, чем на исходных данных, поэтому можно сделать вывод, что факторный анализ не подходит для данной задачи классификации фенотипов заболевания ЖКТ.

Чтобы улучшить результаты логистической регрессии, нужно выполнить детальный анализ групп, чтобы сократить количество исходных признаков и проверить работу логистической регрессии на уменьшенном наборе данных. Возможно, есть признаки, которые негативно влияют на классификации, и, удалив их, можно добиться большей точности на представленных моделях.

2.2. Детальный анализ групп признаков

Проблема отбора значимых признаков для анализа данных заключается в том, что мы не знаем, на основе какого критерия нужно убирать из рассмотрения определённые признаки, потому что для этого нужно обладать дополнительными знаниями в области медицины. Поэтому мы можем судить только на основе следующего критерия: необходимо с помощью экспериментов над различными группами показателей пациентов определить, какую группу признаков можно полностью убрать из рассмотрения ввиду низких результатов классификации, а какую выборку – выбрать для дальнейших исследований.

Для начала следует разбить выборку на различные группы по признакам:

- Во-первых, проведём классификацию по одиночным группам признаков: Р (пищевые привычки), S (уровень стресса), В (симптомы болезни), К (уровень качества жизни), М (лабораторные маркеры).
- Во-вторых, выполним логистическую регрессию по смешанным группам признаков.

В ходе анализа отдельных одиночных групп вопросов получили то, что группа показателей «Симптомы болезни» оказывает весомый вклад при классификации, потому что для этой группы признаков точность логистической регрессии на тестовой выборке составляет 68,97 %. Этот результат намного выше относительно остальных групп признаков, что неудивительно, поскольку объективно симптомы болезни часто хорошо описывают заболевание. Также можно сказать, что группа вопросов «Пищевые привычки» является наименее пригодной для исследования, потому что на ней самые низкие результаты логистической регрессии – 41,38 %, и эта группа вопросов составляет 70 признаков из 112, что больше половины.

На втором этапе анализа групп рассматриваются различные комбинации групп признаков, чтобы определить группу вопросов с наилучшими результатами логистической регрессии и максимальным количеством признаков, для того чтобы использовать этот набор данных для других моделей машинного обучения.

Получили, что точность логистической регрессии на тестовом наборе для всех 112 признаков равняется 51,72 %, что среди остальных смешанных групп вопросов является наименьшим результатом. Учитывая, что результаты среди остальных смешанных групп кардинально не отличаются (рис. 3), то в качестве основной выборки для обучения дальнейших моделей была выбрана смешанная группа SBKM из 42 признаков – именно эта группа содержит все признаки четырёх групп за исключением группы «Пищевые привычки».

Таким образом, на данном этапе мы не используем факторы, а будем выполнять дальнейшие исследования на исходных признаках без группы показателей «Пищевые привычки» по следующим причинам:

- Для группы SBKM точность равна 68,96 %, что является лучшим результатом среди всех групп для логистической регрессии. Таким образом, без группы признаков «Пищевые привычки» точность модели хуже не стала.

```
Train_size = 0.9
Выборка - X_ALL
Best params: {'C': 0.1, 'max_iter': 100, 'penalty': 'l1', 'solver': 'saga'}
Train score: 0.6428571428571429
Test score: 0.5172413793103449

Train_size = 0.9
Выборка - X_SBKM
Best params: {'C': 1, 'max_iter': 1000, 'penalty': 'l1', 'solver': 'saga'}
Train score: 0.75
Test score: 0.6896551724137931

Train_size = 0.9
Выборка - X_BKM
Best params: {'C': 0.1, 'max_iter': 100, 'penalty': 'l2', 'solver': 'newton-cg'}
Train score: 0.6746031746031746
Test score: 0.6896551724137931

Train_size = 0.9
Выборка - X_SBK
Best params: {'C': 0.1, 'max_iter': 100, 'penalty': 'l2', 'solver': 'newton-cg'}
Train score: 0.7023809523809523
Test score: 0.6551724137931034
```

Рис. 3. Результаты логистической регрессии на смешанных группах признаков

- Если более детально изучать признаки в группе «Пищевые привычки», то в ней есть много субъективных вопросов, достоверность ответов на которые неизвестна. Также непонятно, могут ли эти признаки влиять на фенотипы заболевания или нет. Например, среди признаков есть вопросы по типу: «довольны ли Вы разнообразием потребляемых продуктов питания?» или «сколько раз в день Вы пьёте кофе?»
- Определить, какие признаки значимы внутри группы, мы не можем, потому что этим должен заниматься специалист в области медицины.

Здесь же стоит сказать, что кластеризация методом k -средних показала низкие результаты. Точность составила 23,49 % на группе признаков SBKM – это один из худших показателей среди всех моделей (рис. 4). Можно сказать, что точность данной модели является вероятностью случайного угадывания.

Также и для других групп признаков точность кластеризации составляет от 10 до 24 %. Следовательно, для исходной задачи нужно искать более сложную модель.

3. Применение более сложных моделей машинного обучения

После того как были исследованы признаки, проведён факторный анализ и использованы простые модели для классификации, нужно приступить к поиску других, более сложных моделей. Следующей моделью стало решающее дерево.

```
execute_clustering(X_SBKM, y)

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
  warnings.warn(
[0, 0, 0, 0, 0, 0, 3, 0, 5, 5, 5, 4, 4, 0, 5, 4, 5, 5, 3, 3, 0, 0, 0, 2
[0, 0, 0, 0, 0, 0, 4, 2, 1, 5, 1, 5, 3, 3, 4, 5, 3, 3, 1, 1, 0, 3, 2, 5
Adjusted Rand Index: 0.04326588390810449
Accuracy: 23.49%
```

Рис. 4. Результаты кластеризации методом k -средних на наборе SBKM

3.1. Решающее дерево

Решающее дерево представляет такую структуру, что в каждой его вершине находится некоторый вопрос, а рёбра дерева соответствуют вариантам ответов на этот вопрос. В случае бинарного ответа: да или нет. Если в качестве признака выступает числовое значение, то определяется оптимальное пороговое значение, которое разделяет данные на две группы [3].

Решающее дерево даже при большой своей глубине не показывает результата лучше относительно логистической регрессии, несмотря на то, что использовалась кросс-валидация и подбор гиперпараметров по сетке.

Точность составила 62,07 % на тестовом наборе (рис. 5).

```
Best params: {'max_depth': 18, 'min_samples_leaf': 4, 'min_samples_split': 10}
Train score: 0.8253968253968254
Test score: 0.6206896551724138
```

Рис. 5. Точность и лучшие параметры модели решающего дерева

3.2. Случайный лес

Как правило, на практике в машинном обучении более интересной и плодотворной идеей является использование не одного дерева, а комбинации нескольких решающих деревьев. Особенность случайного леса заключается в том, что он генерирует несколько случайных деревьев, каждое из которых обучается на случайном наборе признаков и случайной подвыборке наблюдений из обучающего набора данных. При выдаче ответа алгоритм придерживается мнения большинства деревьев.

Случайный лес является одной из популярных реализаций алгоритма ансамблей деревьев в scikit-learn [4]. Обычно эта модель хорошо работает с небольшими и средними объёмами данных. Случайный лес может эффективно работать даже при наличии нескольких сотен признаков.

Данный алгоритм для рассматриваемой группы признаков даёт результат лучше относительно логистической регрессии, и точность на тестовом наборе составляет 72,41 % (рис. 6).

```
Best params: {'max_depth': 6, 'min_samples_leaf': 4, 'min_samples_split': 10, 'n_estimators': 200}
Train score: 0.7777777777777778
Test score: 0.7241379310344828
```

Рис. 6. Точность и лучшие параметры для случайного леса

3.3. Градиентный бустинг

Случайный лес – это очень мощный алгоритм машинного обучения, но есть более продвинутый алгоритм построения ансамбля решающих деревьев – *бустинг*. Бустинг также строит множество деревьев, но в этом алгоритме построение деревьев происходит последовательно, так что каждое следующее дерево старается по максимуму исправить ошибки всей предыдущей совокупности деревьев [3].

Выполнение `GradientBoostingClassifier` из пакета `scikit-learn` может выполняться на протяжении нескольких часов, что довольно долго при подборе параметров, и в итоге эта модель даёт небольшие результаты – 59 % точности.

Результат является осмысленным, потому что это мощный алгоритм, и он лучше подходит для задач с большим датасетом, который насчитывает десятки и сотни тысяч наблюдений.

3.4. XGBoost

Далее была найдена реализация алгоритма бустинга, которая отличается от предыдущей оптимизацией, позволяющей выполнять обучение и тестирование модели намного быстрее.

Результаты стали хуже относительно случайного леса – 62,06 % (рис. 7), и это также можно объяснить тем, что модель лучше подходит для большого массива данных. Градиентный бустинг, особенно его более продвинутые варианты (например, XGBoost, LightGBM), часто требуют большего объёма данных для достижения высокой точности.

```
Best params: {'learning_rate': 0.01, 'max_depth': 2, 'n_estimators': 200}
Train score: 0.6785714285714286
Test score: 0.6206896551724138
```

Рис. 7. Результаты и параметры XGBoost

3.5. MLPClassifier

Многослойный перцептрон (MLP) представляет собой класс искусственных нейронных сетей с полносвязным типом связи нейронов [5]. Напомним, что нейронная сеть – это последовательность искусственных нейронов, связанных друг с другом. `MLPClassifier` из пакета `scikit-learn` является реализацией такой сложной структуры, как многослойный перцептрон, но в то же время является простым в использовании и довольно быстрым алгоритмом по времени обучения.

Данную модель приходится учить и подбирать параметры для неё вручную, потому что в случае использования валидации модели по сетке параметров лучшая модель будет выбираться по точности на тренировочном наборе, а это не является показателем лучшей модели для данной задачи, так как нас интересует лучший результат на тестовом наборе, который нейросеть ещё не видела.

Проблема связана с тем, что, в отличие от предыдущих моделей, нейронные сети чаще склонны к переобучению, особенно на небольших датасетах, и высокий показатель на тренировочном наборе не гарантирует, что будет высокий показатель на тестовом наборе. Это связано с тем, что модель может просто запоминать тренировочный датасет и не выявлять характерные признаки данных.

Множество экспериментов показало, что лучше всего для данных подходит стохастический градиентный спуск «sgd», количество скрытых слоёв равно трём и количество нейронов (42, 50, 6) соответственно, $\alpha = 0,1$ – коэффициент регуляризации и максимальное количество итераций – 1000 (рис. 8).

```
# Создание и обучение модели
model = MLPClassifier(solver='sgd', hidden_layer_sizes=(42, 50, 6), alpha = 0.1, max_iter=1000)
model.fit(X_train, y_train)
```

Рис. 8. Параметры модели MLPClassifier

В ходе подбора гиперпараметров было лучшим результатом 79,31 % точности на тестовом наборе (рис. 9).

```
MLPClassifier
Accuracy train: 0.8174603174603174
Accuracy test: 0.7931034482758621
```

Рис. 9. Результаты модели MLPClassifier

Это неплохой результат, но больше достигнуть не удалось, потому что при добавлении дополнительных слоёв нейронов сеть быстро переобучается и выдаёт низкие результаты на тестовом наборе.

3.6. Нейронная сеть с использованием Keras

Минусом работы с MLPClassifier является то, что нет возможности гибко задавать параметры между слоями, что является проблемой в обучении. Например, тяжело найти момент во время обучения, когда эта модель переобучается.

В Python есть возможность реализовать нейросеть с гибкой настройкой параметров и слоёв нейронной сети. Для этого можно воспользоваться библиотекой Keras [6].

Однако при помощи данной библиотеки с тонкой настройкой параметров не удалось повысить точность: либо модель сильно переобучается, и тогда точность на тестовом наборе меньше 50 %, либо получается точность, сопоставимая с логистической регрессией, например, для сети с 4 слоями, количеством нейронов – (36,

128, 64, 6) и с использованием дропаута, который является методом регуляризации, точность равна 68,97 % (рис. 10).

```
Test loss TRAIN: 0.730056881904602
Test accuracy TRAIN: 0.7182539701461792
1/1 [=====] - 0s
Test loss TEST: 0.8424545526504517
Test accuracy TEST: 0.6896551847457886
```

Рис. 10. Результаты нейросети, написанной с помощью Keras

3.7. CatBoost

После подбора параметров нейросети было решено вернуться к поиску алгоритмов с деревьями. В какой-то момент была найдена модель градиентного бустинга от «Яндекса» – CatBoost [7].

CatBoost позволяет использовать категориальные признаки без необходимости их предварительно обрабатывать. Данная модель хорошо подходит для небольшого набора данных, что важно в данной задаче. Также преимуществом является возможность обеспечивать повышенную точность за счёт уменьшения переобучения и проводить обучение на нескольких GPU.

Данная модель среди всех показывает лучший результат по точности: 92,85 % на тренировочных данных и 79,31 % правильно определённых фенотипов заболевания ЖКТ на тестовых (рис. 11). Лучший показатель был достигнут при следующих параметрах: глубина деревьев – 12, количество итераций – 200, темп обучения – 0,01.

```
Best params: {'depth': 12, 'iterations': 200, 'learning_rate': 0.01}
Train score: 0.9285714285714286
Test score: 0.7931034482758621
```

Рис. 11. Результаты и параметры модели CatBoost

Если посмотреть на дополнительные метрики *precision* и *recall*, то данный результат является довольно-таки высоким. *Precision* (точность) – процент представителей одного класса среди тех, которых алгоритм отнёс к этому классу. *Recall* (полнота) – процент найденных объектов одного класса среди всех объектов этого класса.

Например, если обратить внимание на полноту по классам на тестовой выборке (рис. 12), то для двух классов модель определяет все наблюдения, для двух других около 89 %, но есть один класс под номером «1» (ПИ-СРК), для которого полнота равна нулю. Детально проанализировав предсказанные значения, было выявлено, что в тестовую выборку попадают 2 представителя первого класса и оба определяются неправильно, отсюда полнота равна нулю. Это может указывать, что есть проблема с разделением наблюдений по классам.

```

Precision для каждого класса TRAIN: Precision для каждого класса TEST:
Класс 0: 1.00                      Класс 0: 1.00
Класс 1: 1.00                      Класс 1: 0.00
Класс 2: 0.96                      Класс 2: 0.50
Класс 3: 0.91                      Класс 3: 0.73
Класс 4: 0.82                      Класс 4: 0.75
Класс 5: 1.00                      Класс 5: 1.00
Recall для каждого класса TRAIN:   Recall для каждого класса TEST:
Класс 0: 1.00                      Класс 0: 1.00
Класс 1: 0.69                      Класс 1: 0.00
Класс 2: 1.00                      Класс 2: 1.00
Класс 3: 0.98                      Класс 3: 0.89
Класс 4: 0.93                      Класс 4: 0.75
Класс 5: 0.93                      Класс 5: 0.88

```

Рис. 12. Precision и recall для модели CatBoost

3.8. Результаты работы моделей

Таким образом, получили следующие результаты на группе из 42 признаков (табл. 1).

Таблица 1. Результаты моделей на группе признаков SBKM, %

Модель	Train	Test
Логистическая регрессия	75,00	68,97
Кластеризация методом k -средних	—	23,49
MLPClassifier	81,74	79,31
Нейросеть Keras	73,00	68,97
Решающее дерево	82,53	62,06
Случайный лес	77,78	72,41
XGBoost	67,86	62,07
CatBoost	92,85	79,31

Для того чтобы убедиться, что лучшие результаты достигаются на группе исходных признаков, состоящей из 4 групп вопросов (без пищевых привычек), покажем результаты на новых факторах (табл. 2) и на всех исходных 112 признаках с группой Р («Пищевые привычки») (табл. 3).

Заключение

Данная работа посвящена поиску модели машинного обучения, которая решает задачу классификации фенотипов заболевания ЖКТ. Были рассмотрены модели и методы машинного обучения: нейронные сети, решающие деревья, логистическая регрессия, факторный анализ, кластерный анализ.

Таблица 2. Результаты моделей на новых факторах, %

Модель	Train	Test
Логистическая регрессия	55,95	44,82
Кластеризация методом k -средних	—	18,86
MLPClassifier	73,80	58,62
Решающее дерево	55,55	58,62
Случайный лес	89,28	51,72
XGBoost	100	48,27

Таблица 3. Результаты моделей на всех исходных признаках, %

Модель	Train	Test
Логистическая регрессия	64,28	51,72
Кластеризация методом k -средних	—	11,39
MLPClassifier	99,20	44,82
Решающее дерево	56,34	44,82
Случайный лес	80,15	68,96
XGBoost	86,11	55,17

Выявлено, что классифицировать фенотипы заболеваний по данной выборке можно, но с разной точностью. Была найдена модель бустинга CatBoost, которая является продвинутой реализацией градиентного бустинга, разработанного в «Яндексе». CatBoost для исходного набора данных на 42 признаках лучше всего подходит для данной задачи и выдаёт следующие лучшие результаты: 92,85 % точности на тренировочной выборке и на тестовой выборке – 79,31 %. Данный результат можно считать неплохим, учитывая проблемы с малым количеством данных, а также рассмотрев дополнительные метрики *precision* и *recall*.

Можно предположить, что в исходном наборе есть аномальные данные, потому что было применено множество моделей, а также для каждой из них был выполнен поиск наилучших параметров, но в итоге точность так и не превысила 90 % на тестовом наборе для всех рассмотренных моделей. В таком случае остаётся искать выбросы в датасете, чтобы повысить точность, но данная задача выходит за рамки данной статьи.

Для дальнейшего продолжения исследования в данной области потребуется собрать больше данных, а также необходим специалист в области медицины, чтобы уточнить достоверность признаков, а также из групп вопросов отобрать признаки, которые с медицинской точки зрения должны вносить большой вклад в диагностику фенотипов заболевания ЖКТ.

Код программ был написан в Google Colab на языке Python. Для создания моде-

лей использовались компоненты библиотеки scikit-learn и Keras. Медицинские данные для исследования были предоставлены Омским государственным медицинским университетом. Результаты работы изображены на графиках, в таблицах и записаны в блокноты Google Colab.

Литература

1. Введение в машинное обучение. URL: <https://habr.com/ru/post/448892/> (дата обращения: 25.02.2024).
2. Агалаков С.А. Статистические методы анализа данных: учеб. пособие. Омск: Изд-во Ом. гос. ун-та, 2017.
3. Быстрый старт в искусственный интеллект. URL: <https://stepik.org/course/80782/> (дата обращения: 31.05.2023).
4. Официальная документация Scikit-learn. URL: <https://scikit-learn.org/stable/index.html> (дата обращения: 25.02.2024).
5. Николенко С., Кадури А., Архангельская Е. Глубокое обучение. СПб.: Питер, 2018.
6. Официальная документация Keras. URL: <https://keras.io/api/> (дата обращения: 25.02.2024).
7. Официальная документация CatBoost. URL: <https://catboost.ai/en/docs/> (дата обращения: 25.02.2024).

APPLICATION OF MACHINE LEARNING MODELS FOR THE DIAGNOSIS OF GASTROINTESTINAL DISEASE PHENOTYPES

A.A. Berezin¹

Ph.D. Student, e-mail: andreyberezin55@gmail.com

S.A. Agalakov²

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

¹Omsk State Technical University, Omsk, Russia

²Dostoevsky Omsk State University, Omsk, Russia

Abstract. The work is devoted to the task of diagnosing the phenotypes of diseases of the gastrointestinal tract using machine learning models. The main purpose of this work is to find a model that solves the problem of classifying patients in a medical institution. In the course of the work, simple models such as logistic regression and cluster analysis were first considered. Due to the inefficiency of these methods, a search was made for more complex models and methods of data analysis: decision trees, ensembles of trees, artificial neural networks. Work was also carried out on the preliminary processing of the data set and the selection of significant features for classification. As a result of the work, the best model for the data set under consideration was found – CatBoost, which diagnoses phenotypes with an accuracy of 92.85 % in the training sample and 79.31 % in the test sample. In addition, the paper evaluates the applicability of machine learning methods in clinical practice and identifies areas for future research for more accurate diagnosis.

Keywords: machine learning, classification problem, logistic regression, neural networks, ensembles of decision trees, boosting.

Дата поступления в редакцию: 29.02.2024

АНАЛИЗ МЕТОДОВ КЛАСТЕРИЗАЦИИ ДЛЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В РОЗНИЧНОЙ ТОРГОВЛЕ

Т.С. Катермина

к.т.н., e-mail: nggu-lib@mail.ru

А.А. Макамбеджан

магистрант, e-mail: aygul-mak@mail.ru

Нижевартовский государственный университет, Нижневартовск, Россия

Аннотация. В статье рассматриваются теоретические основы интеллектуального анализа данных, изучаются методы кластерного анализа на примере выполнения задачи кластеризации данных о продажах от компании «1С». Рассмотренные методы применимы к задаче поддержки принятия решений в различных сферах, в том числе в сфере розничной торговли. В статье приводится обзор методов кластерного анализа, а также эксперимент с применением агломеративного метода кластерного анализа и метода k -средних. Выявлены преимущества и недостатки указанных методов, приведены результаты моделирования.

Ключевые слова: искусственный интеллект, кластерный анализ, интеллектуальный анализ данных.

Введение

Интеллектуальный анализ данных (далее – ИАД) – это процесс поддержки принятия решений, основанный на поиске в данных скрытых закономерностей (шаблонов информации) [1]. Для обнаружения скрытых знаний применяют специальные методы автоматического анализа – DataMining. К наиболее известным методам представления и получения знаний можно отнести классификацию, регрессию, кластеризацию, прогнозирование временных рядов и т. п. [2]. К задачам, с решением которых может помочь ИАД, относятся:

- 1) получение отчётов;
- 2) решение нестандартных вопросов (например, классификация);
- 3) построение моделей и прогнозирование.

Область применения ИАД обширна и включает в себя все сферы общества: научную, социальную, экономическую и политическую.

К одному из самых востребованных методов интеллектуального анализа данных относится кластеризация.

Для выполнения задачи разделения группы товаров на категории по соотношению цены и качества можно прибегнуть к использованию кластеризации. Под термином «кластеризация» понимается группировка объектов на основе близости их свойств. Группы, по которым распределяются данные, называются кластерами.

Каждый из них состоит из схожих объектов, но элементы разных кластеров должны отличаться друг от друга.

Иными словами, кластеризация – процедура, которая любому объекту $x \in \mathbb{X}$ ставит в соответствие метку кластера $y \in \mathbb{Y}$, из чего можно сделать вывод, что основной задачей кластеризации является разбиение векторов на подмножества, которое ориентируется на выявленные сходства и различия элементов [3].

Кластеризация обычно применяется при отсутствии априорной информации о группах, по которым будут классифицироваться данные, или когда кластеризируемых объектов слишком много для ручного анализа.

В задачах кластеризации нет необходимости в указании выходной переменной, а число кластеров может быть неизвестным. Стоит отметить, что кластеризация указывает только на схожесть объектов и не даёт готовый ответ в процессе решения задачи. Например, не указывает на закономерности, по которым были сформированы кластеры. Кластеризация участвует в решении следующих задач:

1. Изучение данных. Кластерный анализ данных помогает выявить структуру данных, сделать их наглядными и удобными для восприятия человека. При помощи кластерного анализа можно выявить необходимые параметры для дальнейшего моделирования сложных систем [4].
2. Облегчение анализа. Кластеризация упрощает последующую обработку данных и построение моделей, также появляется возможность создать индивидуальную модель каждого кластера.
3. Сжатие данных. Кластеризация может помочь сохранить объём хранимых данных, сохранив по одному наиболее приближённом к остальным представителю от каждой группы.
4. Прогнозирование. Кластерный анализ может определить, в какой кластер будет отнесён новый объект исходя из критериев кластеризации, и предсказать его поведение, опираясь на информацию о схожих элементах.
5. Обнаружение аномалий. Кластеризация способна определять нетипичные объекты, которые нельзя отнести ни к одному из кластеров.

Проводить кластеризацию можно по нескольким признакам, после чего будет достаточно просто определить, объектов с какими характеристиками больше, а каких – меньше, и сделать выводы из получившихся результатов. Кластеризация применяется в:

- 1) маркетинге;
- 2) сегментации изображений;
- 3) медицине;
- 4) археологии;
- 5) социологии;
- 6) государственном управлении и т. д.

Методы кластерного анализа

В связи с тем, что одно и то же множество объектов можно разделить на кластеры по-разному, появилось большое количество методов кластеризации. Не существует одного универсального метода кластеризации, но можно подобрать из них

наиболее подходящий под ту или иную ситуацию.

Однако, существуют сложности, с которыми предстоит столкнуться при использовании кластерного анализа:

1. Неопределённость в выборе критерия качества кластеризации. Проблема заключается в том, что на практике, когда объекты описываются десятками признаков, определение их расположения становится затруднительным.
2. Трудность выбора меры близости. Она обусловлена различной природой данных. Например, если для вычисления расстояния между числовыми данными можно использовать евклидово расстояние, то категориальным типам подобные меры не подойдут. Нужно будет прибегать к использованию специальной меры, которая, например, задаётся функцией отличия.
3. Различные требуемые машинные ресурсы. Обычно, чем точнее результат выдаёт кластер, тем больше он потребует затраченного времени и памяти. Поэтому в интеллектуальном анализе не получил распространение иерархический метод кластеризации, который строит полное дерево вложенных кластеров.

Классификация методов кластерного анализа по способу обработки данных:

1. Иерархические методы:
 - агломеративный метод AGNES (Agglomerative Nesting);
 - дивизимный метод DIANA (Divisive Analysis).
2. Неиерархические методы:
 - метод k -средних (k -means);
 - метод разделения вокруг метоидов PAM (partitioning around medoids);
 - метод кластеризации категориальных данных CLOPE и т. д.

Также кластеризацию можно поделить на чёткую (например, метод k -средних) и нечёткую (например, метод c -средних).

При иерархической кластеризации выполняется последовательное объединение меньших кластеров в большие или разделение больших кластеров на меньшие. Она предоставляет достаточно точный результат, так как вырисовывает дерево вложенных кластеров, но при этом является более вычислительно дорогостоящей, так что её не рекомендуется использовать для кластеризации больших наборов данных. Пример дерева вложенных кластеров приведён на рис. 1.

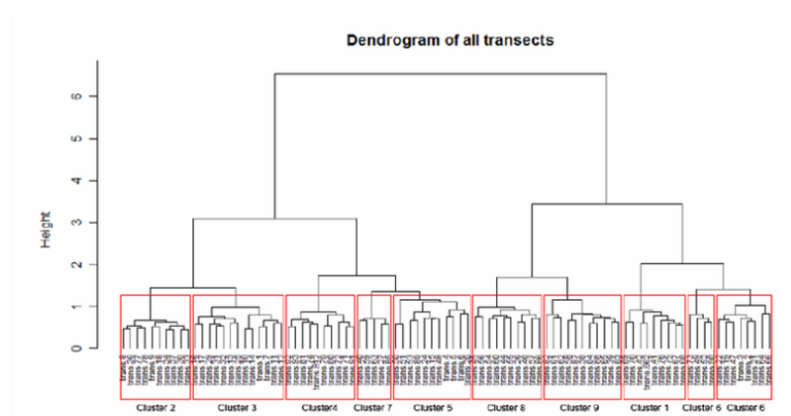


Рис. 1. Дендрограмма иерархической кластеризации [5]

Агломеративный иерархический метод характеризуется последовательным объединением исходных объектов и соответствующим уменьшением числа кластеров. На начальном этапе работы алгоритма все элементы являются кластерами. На первом шаге объекты, которые имеют схожие характеристики, объединяются в кластер. Далее объединение продолжается до тех пор, пока все объекты не будут составлять один кластер.

Дивизимный иерархический метод подразумевает собой последовательное разделение исходного кластера, состоящего из всех элементов, на кластеры меньшего размера, в результате чего образуется последовательность расщепляющих групп.

Метод k -средних – наиболее популярный метод кластерного анализа. Он стремится минимизировать суммарное квадратичное отклонение точек кластеров от центров этих кластеров. Данный метод разбивает множество элементов векторного пространства на заранее известное количество кластеров k . Суть метода k -средних такова: на каждой итерации переисчисляется центр масс для каждого кластера, который был получен на предыдущем шаге, затем векторы снова разбиваются на кластеры в соответствии с тем, какой из центров оказался ближе по выбранной метрике.

Пример результата кластеризации приведён на рис. 2.

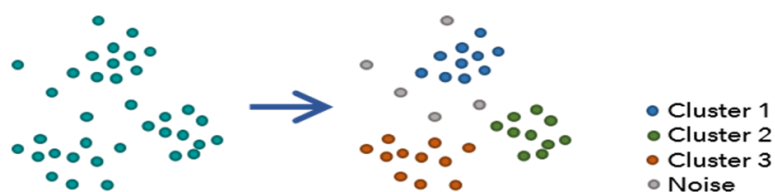


Рис. 2. Кластеризация на основе плотности [6]

Сравнительный анализ методов кластерного анализа будет проводиться на примере датасета Predict Future Sales (<https://www.kaggle.com/competitions/competitive-data-science-predict-future-sales/data>). Этот набор данных содержит ежедневные данные о продажах, предоставленные одной из крупнейших российских компаний – разработчиков программного обеспечения – фирмой «1С» – за период с 1 января 2013 г. по 31 октября 2015 г.

Набор данных состоит из 6 csv-файлов:

- `item_categories.csv`, который содержит в себе информацию о категориях товаров;
- `items.csv`, который включает в себя сведения о самих товарах;
- `sales_train.csv`, для обучения модели;
- `test.csv`, который будет проверочным набором для модели;
- `sample_submission.csv`, который является шаблоном для вывода результата.

Содержимое файла `sales_train.csv`, над данными из которого проводился кластерный анализ, представлено в табл. 1.

В данной работе были рассмотрены следующие методы кластеризации: агломеративная иерархическая кластеризация и метод k -средних.

Таблица 1. Содержимое файла для обучения модели

<i>date</i>	<i>datenum</i>	<i>shopid</i>	<i>itemid</i>	<i>itemprice</i>	<i>itemcntday</i>
02.01.2013	0	59	22154	999	1
03.01.2013	0	25	2552	899	1
06.01.2013	0	25	2554	1709.05	1
15.01.2013	0	25	2555	1099	1
10.01.2013	0	25	2564	349	1
02.01.2013	0	25	2565	549	1
04.01.2013	0	25	2572	239	1
11.01.2013	0	25	2572	299	1

С одной стороны, иерархическая кластеризация способна продемонстрировать дерево кластеров, что значительно упростит пользователю выбор количества необходимых кластеров, так как визуализированную информацию человеку проще воспринять. Но, с другой стороны, данное преимущество имеет один фатальный недостаток – при работе с большим объёмом данных процесс кластеризации длится слишком долго. Метод k -средних считается одним из самых быстрых, доступных на данный момент и достаточно прост в реализации.

Класс `KMeans` реализован в библиотеке `sklearn` (`scikit-learn.org`). Визуализацию результатов кластеризации берёт на себя менеджер графического интерфейса фигур `ruplot` из библиотеки `matplotlib` (`matplotlib.org`). Он обеспечивает неявный способ построения графиков подобно `MATLAB`. Графический интерфейс `ruplot` и пример результата кластеризации проданной компанией «1С» продукции за 2 января 2013 г. продемонстрированы на рис. 3.

Класс `AgglomerativeClustering`, как и `KMeans`, находится в библиотеке `sklearn` (`scikit-learn.org`). Результат агломеративной иерархической кластеризации представлен на рис. 4.

Кластерный анализ проводился по столбцам `itemprice` (цена товара) и `itemcntday` (количество проданных товаров).

За 2 января 2013 г. компания «1С» продала 6 684 предмета, которые были распределены двумя методами кластерного анализа по 5 кластерам.

Метод k -средних распределяет товары из набора данных в среднем за 0,06383 секунды следующим образом:

- 1 кластер – 55 элементов;
- 2 кластер – 4 029 элементов;
- 3 кластер – 48 элементов;
- 4 кластер – 1 661 элемент;
- 5 кластер – 891 элемент.

С другой стороны, агломеративный иерархический метод затрачивает в среднем 0,73307 секунды и даёт следующие результаты:

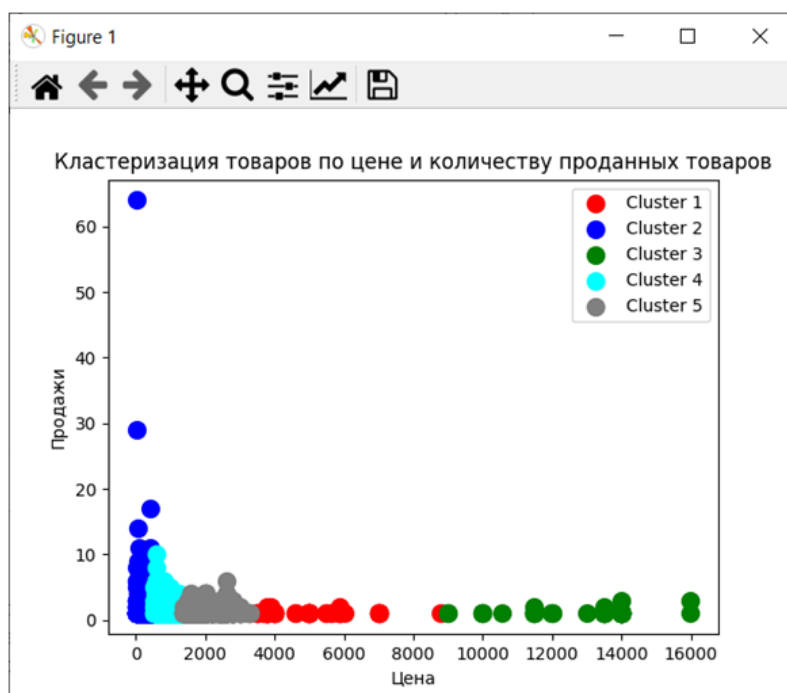


Рис. 3. Результат кластеризации всей продукции компании «1С» за 2 января 2013 г. методом KMeans

- 1 кластер – 855 элементов;
- 2 кластер – 49 элементов;
- 3 кластер – 4 930 элементов;
- 4 кластер – 28 элементов;
- 5 кластер – 822 элемента.

Как видно, ни один кластер одного метода кластерного анализа не соответствует какому-либо кластеру другого метода.

Ориентируясь на полученные результаты, можно подчеркнуть, что самым прибыльным кластером, выделенным агломеративным методом, является 1, а самым неприбыльным – 4. В то же время самый прибыльный кластер, определённый методом k -средних, – 5, а тот, что принёс меньше всего прибыли, – 1.

Также можно сделать вывод, что разные методы кластеризации дают разные результаты. Однако расположение и содержание некоторых кластеров может примерно совпадать, как, например, у 2 кластера по методу агломеративной кластеризации и 3 кластера по методу k -средних. Но дать точный ответ на вопрос: «Какой метод кластерного анализа из представленных справился лучше?» – может только специалист в данной области применения анализа данных, а точнее в области розничной торговли, поэтому основные параметры, по которым будут сравниваться алгоритмы в данной работе, – это скорость и удобство использования.

Если ориентироваться на скорость выполнения, метод агломеративной иерархической кластеризации обработал 6 684 элементов в 11,5 раз медленнее, чем метод k -средних, что значительно повлияет на комфортность и эффективность работы менеджера магазина розничной торговли. Для специалиста очень важно быстро прове-

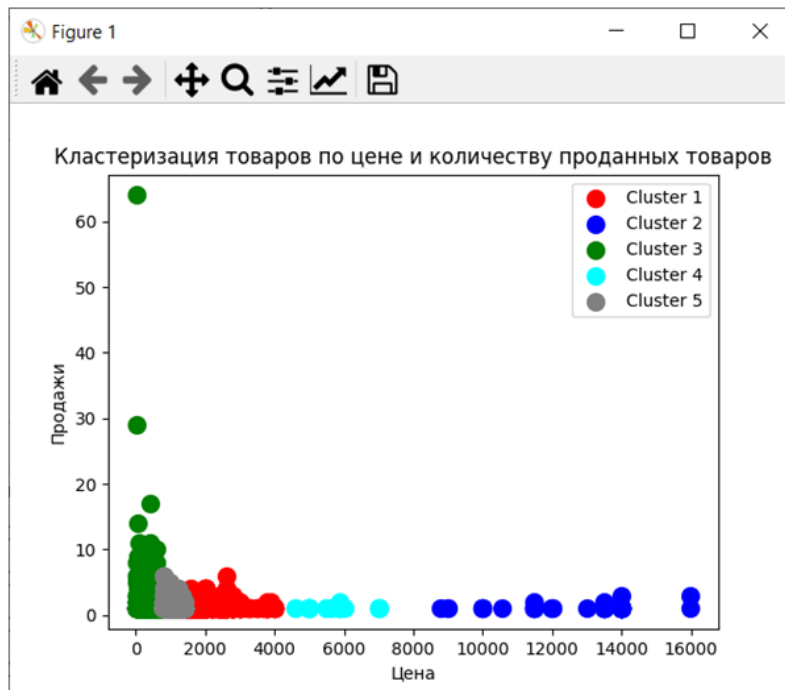


Рис. 4. Результат кластеризации всей продукции компании «1С» за 2 января 2013 г. методом AgglomerativeClustering

сти анализ сведений о проданной продукции и принять наиболее выгодное решение исходя из полученных результатов. Также в рамках анализа был проведён кластерный анализ 63 170 элементов (продажи компании «1С» за январь 2013 г.) методом k -средних. Кластеризация была проведена за 0,37588 секунды, что в 2 раза быстрее, чем кластеризация агломеративным методом набора данных из 6 684 элементов.

Заключение

Интеллектуальный анализ данных может быть использован в самых разных областях: от медицины и до экономики. Являясь одним из самых популярных методов ИАД, кластерный анализ способен решить задачу разделения данных на кластеры по некоторым характеристикам. Однако разные методы кластерного анализа обладают разной эффективностью и спецификой. Так метод k -средних в большей степени подходит для решения задачи кластеризации больших объёмов данных, с которыми иногда приходится иметь дело в решении задач в области розничной торговли. Он не требует такого большого количества временных ресурсов, как агломеративный иерархический метод, а значит, способен за то же время провести анализ большего количества данных, благодаря чему работа менеджера по продажам будет проходить гораздо эффективнее.

Литература

1. Остроух А.В., Николаев А.Б. Интеллектуальные информационные системы и технологии : моногр. СПб.: Лань, 2019.
2. Заболотникова В.С., Ромашкова О.Н. Анализ методов кластеризации для эффективного управления процессами в налоговой службе // *Фундаментальные исследования*. 2017. № 9-2. С. 303–307.
3. Ризаев И.С., Рахал Я. Интеллектуальный анализ данных для поддержки принятия решений. Казань: Школа, 2011.
4. Яковлева Е.А., Катермина Т.С., Шарич Э.Э., Яковлева Д.Д. Формирование потенциала финансовой системы для повышения инновационной активности // *Вопросы инновационной экономики*. 2019. Т. 9, № 2. С. 349–360.
5. Heiss M. The breeding bird communities of the Talish mountains (Azerbaijan) and their response to forest degradation: Diploma thesis. University of Greifswald, 2010. URL: https://www.researchgate.net/figure/The-nine-breeding-bird-communities-of-the-Talish-mountains-represented-by-clusters-The_fig2_274533232 (дата обращения: 10.11.2023).
6. Find Point Clusters (GeoAnalytics). URL: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/big-data-analytics/find-point-clusters.htm> (дата обращения: 10.11.2023).

ANALYSIS OF CLUSTERING TECHNIQUES TO SUPPORT RETAIL DECISION-MAKING

T.S. Katermina

Ph.D. (Techn.), e-mail: nggu-lib@mail.ru

A.A. Makambedzhan

Master's Degree, e-mail: aygul-mak@mail.ru

Nizhnevartovsk State University, Nizhnevartovsk, Russia

Abstract. The article considers theoretical bases of data mining, studies methods of cluster analysis on the example of the task of clustering data on sales from "1С". The methods considered apply to the task of supporting decision-making in various spheres, including retail trade. The article provides an overview of methods of cluster analysis, as well as an experiment with the application of agglomerative method of cluster analysis and *k*-means method. Advantages and disadvantages of these methods are revealed, results of modeling are given.

Keywords: artificial intelligence, cluster analysis, data mining.

Дата поступления в редакцию: 19.11.2023

ГЕНЕТИЧЕСКИЕ АЛГОРИТМЫ ДЛЯ РАЗМЕЩЕНИЯ НАЗЕМНЫХ СТАНЦИЙ В ГЛОБАЛЬНОЙ СПУТНИКОВОЙ СЕТИ СВЯЗИ

Т.В. Леванова^{1,2}

к.ф.-м.н., доцент, e-mail: levanovat@gmail.com

А.В. Гетте²

аспирант, e-mail: gette4aleks@yandex.ru

¹Омский государственный университет им. Ф.М. Достоевского, Омск, Россия

²Институт математики им. С.Л. Соболева СО РАН, Омский филиал, Омск, Россия

Аннотация. Глобальные спутниковые сети обеспечивают связь не только в различных точках Земли, но и в отдалённых регионах, где прокладка кабельных линий невозможна или затруднена. Для обеспечения связи на больших территориях необходимо разместить наземные станции, которые взаимодействуют со спутниковыми системами. Решение проблемы размещения и проектирования наземных станций можно разделить на два этапа. На первом этапе определяют набор строящихся наземных станций и конфигурации их оборудования. На втором этапе происходит регулярный расчёт маршрутов пакетов данных в каждом временном интервале с учётом запросов от клиентов и текущего состояния сети связи. Качество обслуживания клиентов в интервале времени измеряется как среднее количество потерянных или не маршрутизированных пакетов в единицу времени. Ранее при решении задачи маршрутизации предполагалось, что все станции и передающие устройства уже выбраны. В данной работе задача рассматривается как двухэтапная, в которой необходимо решать вопросы первого и второго этапов вместе. Критерием служит минимизация общего числа потерянных пакетов. Для этого построены генетический алгоритм и алгоритм генетического локального поиска, выполнено их экспериментальное исследование, проведено сравнение с другими алгоритмами.

Ключевые слова: задача размещения, задача маршрутизации, генетический алгоритм, локальный поиск, глобальная сеть.

Введение

Глобальные системы связи являются современным средством передачи данных. Они обеспечивают связь в любой точке Земли и широко используются в нефтегазовой отрасли, сетевой торговле, спасателями, путешественниками и т. п. Организация спутниковой связи – это высокотехнологичный процесс, требующий значительных затрат, в том числе на оборудование. Для обеспечения связи на больших территориях необходимо разместить наземные станции, которые будут взаимодействовать со спутниковыми системами.

В данной работе предложены варианты решения задачи определения мест расположения и конфигураций наземных станций в глобальной сети спутниковой связи. Предполагается, что сеть состоит из спутников на низкой околоземной орбите, наземных станций (обеспечивающих доступ в Интернет) и центра управления сетевыми операциями (см., например, [1, 2]). От источника данных пакеты информации передаются через наземные станции спутникам, а затем со спутников – потребителям. В каждом потенциальном месте расположения такая станция может быть создана в соответствии с одним из возможных вариантов или не создаваться вообще. Конструкция наземной станции определяется пропускной способностью соединений со спутниками, пропускной способностью соединений с другими наземными станциями через наземную сеть «Интернет» и стоимостью использования данной конструкции наземной станции в единицу времени. Предполагается, что система работает в дискретных временных интервалах, которые достаточно малы, чтобы предположить, что качество каналов связи и требования клиентов не меняются в течение одного периода.

Маршруты пакетов данных для каждой пары «источник–назначение» рассчитываются в центре управления сетевыми операциями в режиме реального времени, и каждый узел (спутниковая или наземная станция) получает маршруты для всех пакетов, исходящих от этого узла. Каждый пакет, отправляемый из источника в пункт назначения, содержит некоторое содержимое данных и маршрут пакета. Максимально допустимое число рёбер в путях передачи пакетов вводится из-за технического ограничения на количество битов, зарезервированных для кодирования пути передачи пакетов. Короткие пути передачи пакетов также обычно имеют низкую задержку. Для простоты предполагается, что каждая система описывается за один временной интервал, и все требования к этому временному интервалу известны заранее.

Таким образом, для решения проблемы создания глобальной спутниковой связи на первом этапе необходимо решить задачу определения места размещения и выбора конфигурации наземных станций; на втором этапе требуется решить задачу маршрутизации пакетов информации. Решение проблемы маршрутизации пакетов описано в работах [3, 4]. В этих работах все передающие устройства считались уже выбранными и сконструированными. Новизна данной работы заключается в том, что задача определения местоположения и настройки наземных станций для глобальной сети спутниковой связи (задача первого этапа) объединена с задачей маршрутизации пакетов (задача второго этапа). Для решения этой двухэтапной задачи оптимизации предложены генетические алгоритмы, выполнено их экспериментальное исследование с предыдущими разработками.

Работа организована следующим образом. В разделе 1 сформулированы задачи маршрутизации пакетов, размещения и проектирования наземных станций, выписаны математические модели. В разделе 2 описаны разработки авторов для решения задачи размещения наземных станций. В разделе 3 показаны вычислительные исследования, проведён сравнительный анализ результатов алгоритмов решения поставленной задачи.

1. Проблема размещения и проектирования станций

Решение проблемы размещения и проектирования наземных станций можно разделить на два этапа: размещение наземных станций и выбор проектной конфигурации для каждой из них (первый этап); маршрутизация информационных пакетов между клиентами в течение определённого интервала времени (второй этап). Переменные решения на первом этапе определяют набор строящихся наземных станций и варианты их проектирования. На втором этапе рассчитываются маршруты пакетов в каждом временном интервале с учётом сеансовых запросов от клиентов и текущего состояния сети связи. Качество обслуживания в интервале времени измеряется как среднее количество потерянных или не маршрутизированных пакетов в единицу времени, т. е. разница между количеством пакетов для запрошенных сеансов и количеством маршрутизированных пакетов. Количество маршрутизируемых пакетов рассматривается в качестве критерия максимизации. В качестве ограничения вводится общий бюджет на использование наземных станций в течение определённого периода.

1.1. Постановка задачи маршрутизации пакетов

В задаче маршрутизации пакетов даны: $G = (V, E)$ – оргграф, где V – набор спутников и наземных станций, а E – множество соединений (каналов); $n = |V|$, $m = |E|$; $u(e) \geq 0$ – пропускная способность соединения e ; N – общее количество активных сеансов в системе.

Сессия i представлена тройкой (A_i, B_i, W_i) , $i = 1, \dots, N$, где $A_i, B_i \in V$ – пара узлов «источник–назначение», W_i – трафик данных за единицу времени во время сессии. Каждая сессия рассматривается как связь между узлами графа G (наземными станциями или спутниками), к которым в данный момент подключены пользователи; L – максимально допустимое число рёбер в пути передачи пакета; $\tau(e)$ – задержка в передаче данных по каналу $e \in E$.

Задача заключается в том, что требуется найти набор путей в G таким образом, чтобы каждая сессия соответствовала не более чем одному маршруту, и так, чтобы маршрутизировалось максимально возможное количество пакетов.

Задача имеет ограничения, заключающиеся в том, что максимальное количество дуг на каждом пути не превышает L и что для каждой дуги e общий объём информации, передаваемой по всем проходящим маршрутам, не должен превышать пропускную способность $u(e)$ [3, 4].

В предыдущих публикациях теоретически и экспериментально изучались различные подходы к решению проблемы маршрутизации пакетов и обсуждались основанные на графах модели сети с различными уровнями детализации. В этих работах все передающие устройства считались уже выбранными и сконструированными [3, 4]. В настоящей работе, наоборот, решается задача выбора мест размещения и конструкций, а алгоритм решения задачи маршрутизации используется как «чёрный ящик».

1.2. Постановка задачи размещения и проектирования наземных станций

В каждом потенциальном месте расположения наземной станции такая станция может быть создана в соответствии с одним из возможных вариантов или не создаваться вообще. Конфигурация наземной станции характеризуется: 1) пропускной способностью соединений со спутниками; 2) пропускной способностью соединений с другими наземными станциями; 3) стоимостью использования этой станции в единицу времени.

Введём обозначения: I – множество местоположений для наземных станций; D – допустимый бюджет, который будет использован на строительство и эксплуатацию наземных станций; R – набор конфигураций, возможных для каждой наземной станции; c_i^r – стоимость открытия станции конфигурации типа r в пункте i .

Переменные задачи $z_i^r = 1$, если в i размещена станция с конфигурацией r , 0 – иначе, $i \in I$.

Обозначим $f(Z)$ решение проблемы маршрутизации пакетов, заданное матрицей конфигураций $Z = (z_i^r)$. Требуется разместить наземные станции таким образом, чтобы общее количество маршрутизируемых пакетов было максимальным.

Таким образом, математическая модель задачи размещения наземных станций для глобальной сети спутниковой связи будет выглядеть следующим образом:

$$f(Z) \rightarrow \min; \quad (1)$$

$$\sum_{i \in I} \sum_{r \in R} c_i^r z_i^r \leq D; \quad (2)$$

$$\sum_{r \in R} z_i^r \leq 1, i \in I; \quad (3)$$

$$\sum_{i \in I} \sum_{r \in R} z_i^r \geq 1; \quad (4)$$

$$z_i^r \in \{0; 1\}; i \in I, r \in R. \quad (5)$$

Здесь (1) отражает цель выбрать места размещения и оборудование наземных станций, обеспечивающих наименьшее число потерянных пакетов. Условие (2) задаёт ограничение на бюджет. Неравенство (3) позволяет размещать в каждом пункте станцию только с одним типом конфигурации. Выражение (4) учитывает, что для работы всей сети необходимо размещение хотя бы одной станции.

С помощью моделей, подобных (1)–(5), описываются так называемые варианты задачи размещения. В них помимо выбора места размещения необходимо определить вариант функционирования предприятия, тип оборудования, вид продукции и т. п. Информацию о таких задачах можно найти, например, в [5, 6].

2. Алгоритмы решения

В данной работе для рассматриваемой задачи предложены варианты генетического алгоритма и генетического локального поиска.

Генетический алгоритм представляет собой универсальную схему решения широкого круга прикладных задач (см., например, [7–10]). Его основная идея состоит в

компьютерном моделировании эволюционного процесса. Для описания алгоритма традиционно применяются термины из биологии.

В начале работы алгоритма задаётся «искусственная» популяция особей, каждая из которых представляет собой элемент в пространстве решений задачи. Приспособленность особей описывается с помощью некоторой монотонной функции от значения целевой функции задачи. Приспособленность тем выше, чем лучше решение. Как и в биологии, имитируется мутация генов и рекомбинация. Развитие популяции происходит за счёт исключения из неё более «слабых» особей и добавления более «сильных». Отбор осуществляется с помощью различных операторов селекции (пропорциональной, турнирной, срезающей).

Генетический локальный поиск – это метод оптимизации, который сочетает в себе преимущества генетического алгоритма и локального поиска. Основная идея генетического локального поиска заключается в том, что генетический алгоритм используется для создания первоначальной популяции, а затем локальный поиск применяется к каждому из полученных решений. Генетический локальный поиск считается более эффективным, чем простые генетические алгоритмы, потому что он способен находить оптимальные или близкие к оптимальным решения за меньшее количество итераций. Кроме того, генетический локальный поиск более устойчив к проблеме преждевременной сходимости, которая возникает у генетических алгоритмов, когда они останавливаются в локальных оптимумах вместо глобальных. Примером задач, которые могут быть решены с использованием генетического локального поиска, являются задачи транспортной логистики, распределения ресурсов и оптимизации производственных процессов и др. [11, 12].

2.1. Генетический алгоритм

Генетический алгоритм (ГА) основан на так называемой стационарной схеме воспроизводства [10], при которой на каждой итерации генерируется пара новых особей, заменяющих пару «неперспективных» особей в популяции. После заданного числа итераций алгоритм завершает работу, лучшее найденное решение возвращается в качестве результата. Создание пары новых генотипов «потомков» начинается с выбора пары «родительских» особей из текущей популяции с помощью вероятностного оператора селекции. К генотипам выбранных особей применяется оператор кроссинговера, заменяющий часть генов одного генотипа на гены другого. Оба полученных после кроссинговера генотипа подвергаются оператору мутации, при котором некоторое подмножество генов изменяется случайным образом.

Для описания схемы алгоритма введём обозначения: $f(Z)$ – функция приспособленности; λ – количество особей; $B = \{1, \dots, \lambda\}$ – множество всех генотипов, здесь λ – чётное, $x \in B$ – особь; $Sel : B^\lambda \rightarrow \{1, \dots, \lambda\}$ – оператор селекции; $Mut : B \rightarrow B$ – оператор мутации; $Cross : B \times B \rightarrow B \times B$ – оператор кроссинговера. Популяцию поколения $t \geq 0$ обозначим $X^t = (x^{1t}, \dots, x^{\lambda t})$. Итерацией генетического алгоритма является переход от популяции X^t к популяции X^{t+1} . Схема генетического алгоритма приведена ниже (Algorithm 1).

В данной работе при построении генетического алгоритма используются турнирная селекция, одноточечный кроссинговер, одноместный оператор мутации [9].

Algorithm 1 Схема генетического алгоритма

-
- 1: *Инициализация*: положить $t \leftarrow 0$.
 - 2: **for** $k = 1, \dots, \lambda$ **do** построить случайным образом особь $x^{k,0}$.
 - 3: **end for**
 - 4: **while** не выполнено условие остановки **do**
 - 5: **for** $k \leq \lambda/2$ **do** выполнять:
 - 6: *Sel*: выбрать особи $x \leftarrow x^{Sel(X^t),t}, y \leftarrow x^{Sel(X^t),t}$.
 - 7: *Cross*: построить $(x', y') \leftarrow Cross(x, y)$.
 - 8: *Mut*: положить $x^{2k-1,t+1} \leftarrow Mut(x')$, $x^{2k,t+1} \leftarrow Mut(y')$.
 - 9: **end for**
 - 10: Положить $t \leftarrow t + 1$.
 - 11: **end while**
 - 12: *Результат*: лучшая найденная особь.
-

2.2. Генетический локальный поиск

Как и в случае с классическим генетическим алгоритмом, генетический локальный поиск начинается с генерации первоначальной популяции решений, которая затем проходит через процесс селекции, кроссинговера и мутации для формирования новых поколений популяции. Однако, в отличие от классического генетического алгоритма, генетический локальный поиск включает в себя использование локального поиска на каждой итерации для улучшения подмножества решений, отобранных из популяции. Таким образом, генетический локальный поиск объединяет преимущества глобального поиска и локальной оптимизации для достижения более точных решений в задачах оптимизации. Схема алгоритма генетического локального поиска приведена ниже (Algorithm 2).

Algorithm 2 Схема генетического локального поиска

-
- 1: *Инициализация*.
 - 2: Построить начальную популяцию.
 - 3: **while** не выполнено условие остановки **do**
 - 4: Выбрать двух родителей из популяции.
 - 5: Построить по ним новое решение.
 - 6: Применить к нему алгоритм локального улучшения.
 - 7: Выбрать двух родителей из популяции.
 - 8: **if** новое решение лучше худшего в популяции **then**
 - 9: обновить популяцию.
 - 10: **end if**
 - 11: **end while**
 - 12: *Результат*: лучшая найденная особь.
-

2.3. Поиск с чередующимися окрестностями

Для сравнительного анализа работы предложенных алгоритмов использовался алгоритм локального поиска с чередующимися окрестностями (Variable Neighborhood Search Algorithm, VNS). Этот алгоритм, так же как и генетические алгоритмы, принадлежит к методам приближённого решения. Он основан на систематической смене окрестностей [13]. В настоящее время этот подход применяется к широкому кругу прикладных задач (см., например, [13–15]). В данной работе используется вариант алгоритма, предложенный ранее авторами к задаче проектирования глобальной сети спутниковой связи [16]. Для его реализации были введены три вида окрестностей, учитывающих специфику задачи. Одна из них (в [16] и здесь обозначена N_1) использовалась в генетическом локальном поиске.

Сопоставим матрице Z вектор y так, что $y_i = r \Leftrightarrow z_i^r = 1, i \in I, r \in R$.

Определение 1. Окрестностью N_1 целочисленного вектора y называется множество допустимых решений y' , полученных путём увеличения одной из компонент y_i на 1, $i \in I$.

3. Экспериментальные исследования

Тестовые примеры были построены на основе серии тестов для задачи размещения и проектирования наземных станций в глобальной спутниковой сети [4]. Пары узлов были сгенерированы для приближённого моделирования коммуникационных потоков в глобальном масштабе. Предполагалось, что количество пользователей пропорционально численности населения. Исходными и конечными узлами каждого сеанса были случайно сгенерированные пользователи. Пользователи были привязаны к ближайшей наземной станции или спутнику. Предполагается, что пропускная способность канала находится в пределах от 100 до 300 Мбит/с. Скорость передачи данных в каждом сеансе по служебному каналу голосовой связи составляет 9600 бит/с [16].

Вычислительный эксперимент был проведён на 6 тестовых примерах, в каждом из них общее количество переданных пакетов составило 2 355 465 600. Вычисление начального количества потерянных пакетов в каждом примере описано в работах [3, 4], результаты приведены в табл. 1.

Таблица 1. Начальное количество потерянных пакетов

Пример 1	Пример 2	Пример 3	Пример 4	Пример 5	Пример 6
2435×10^5	2235×10^5	2219×10^5	2477×10^5	2220×10^5	2237×10^5

Следует ожидать, что комплекс программ, в котором на первом этапе проводится оптимизация, должен улучшить предыдущие результаты из [3, 4]. Однако было интересно, насколько возможно такое улучшение.

Для каждого примера алгоритмы работали в течение одного и того же промежутка времени, а именно 5, 10, 15, 30, 60 минут. Такая продолжительность считается

приемлемой с практической точки зрения. В генетическом алгоритме использовались турнирная селекция; односточный кроссинговер с вероятностью 0,9; односточный оператор мутации с вероятностью 0,1. На шаге локального улучшения в алгоритме генетического локального поиска использовалась окрестность N_1 .

На рис. 1 и 2 на горизонтальной оси указаны тестовые примеры, по вертикальной оси отражено количество потерянных пакетов.

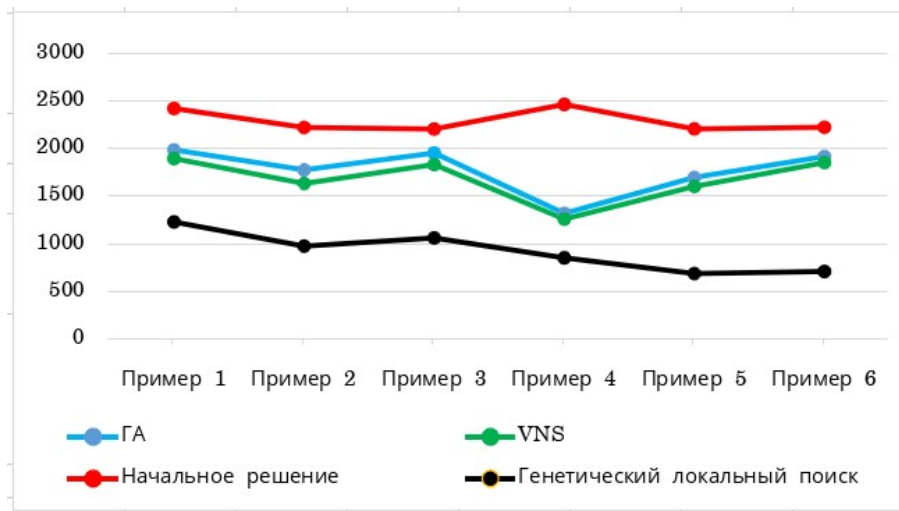
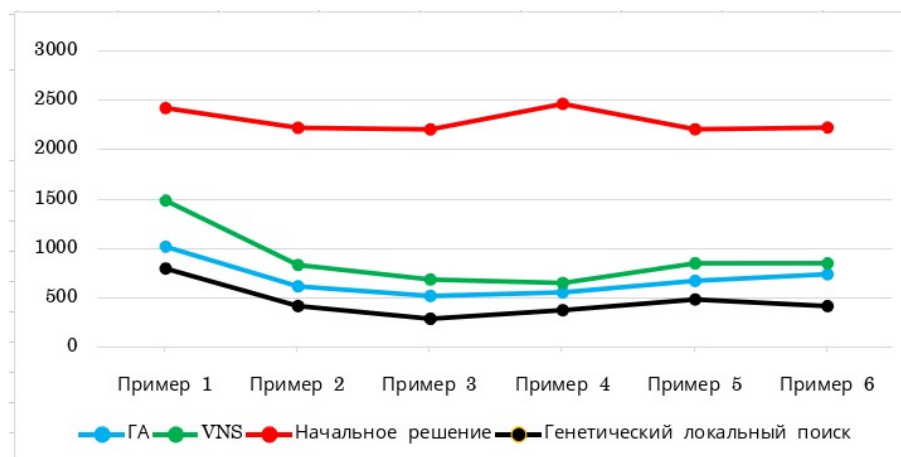


Рис. 1. Среднее количество потерянных пакетов, $t = 15$ мин

В течение первых 5–10 минут тестирования все алгоритмы продемонстрировали примерно одинаковые результаты. Однако генетический локальный поиск несколько выделился за счёт снижения количества потерянных пакетов на 11 % в течение этого времени (см. рис. 1). Изменения начинаются через 15 минут, генетический локальный поиск демонстрирует наименьшее количество потерянных пакетов. По прошествии 60 минут тестирования алгоритмов были получены следующие результаты: генетический локальный поиск увеличил показатели на 85 % в сравнении с генетическим алгоритмом, уменьшившим количество потерянных пакетов на 70 %, и алгоритмом VNS, который показал уменьшение потерь на 60 % (рис. 2).

Заключение

В работе предложены возможные способы решения задачи размещения наземных станций для проектирования глобальной спутниковой сети связи. С этой целью построены проблемно-ориентированные варианты генетического алгоритма и алгоритма генетического локального поиска. Проведено их экспериментальное исследование на примерах, сформированных по аналогии с реальными данными. В результате исследования было установлено, что генетический локальный поиск обеспечивает наименьшее количество потерянных пакетов. Он позволил снизить количество потерь на 85 %, что является значительным улучшением. Генетический алгоритм и локальный поиск с чередующимися окрестностями несколько уступили

Рис. 2. Среднее количество потерянных пакетов, $t = 60$ мин

этим результатам, уменьшив количество потерянных пакетов по сравнению с начальным на 70 и 60 % соответственно. Проведённое исследование имеет практическую значимость для спутниковых операторов и системных инженеров, предоставляющих услуги связи через глобальные спутниковые сети связи. Разработки могут быть использованы в дальнейшем для оптимизации работы указанных сетей связи.

Благодарности

Благодарим А.В. Еремеева и П.А. Борисовского за полезные консультации. Работа выполнена при поддержке гранта РНФ, № 21-41-09017.

Литература

1. Xu S., Wang X.W., Huang M. Software defined next-generation satellite networks: Architecture, challenges, and solutions // IEEE Access. 2018. Vol. 6. P. 4027–4041.
2. Tang Z., Zhao B., Yu W., Feng Z., Wu C. Software defined satellite networks: Benefits and challenges // 2014 IEEE Computers, Communications and IT Applications Conference. 2014. P. 127–132.
3. Borisovsky P., Eremeev A., Hrushev S., Teplyakov V., Vorozhtsov M. On three approaches to length-bounded maximum multicommodity flow with unit edge-lengths // Yugoslav Journal of Operations Research. 2019. Vol. 29, No. 1. P. 93–112.
4. Borisovsky P., Eremeev A., Hrushev S., Teplyakov V. Experimental evaluation of algorithms for packet routing in software defined network // IFAC-PapersOnLine. 2022. Vol. 55, Iss. 10. P. 584–589.
5. Колоколов А.А., Леванова Т.В., Позняков Ю.С. Алгоритмы искусственной иммунной системы для вариантной задачи размещения телекоммуникационных центров // Известия Иркутского государственного университета. Серия: Математика. 2013. Т. 6, № 1. С. 35–44.
6. Levanova T., Gнусарев А. Development of threshold algorithms for a location problem with elastic demand // Lecture Notes in Computer Science. 2018. Vol. 10665 LNCS. P. 382–389.

7. Борисовский П.А. Генетические алгоритмы для задачи о поставках продукции // Динамика систем, механизмов и машин: материалы V Междунар. науч.-техн. конф. Омск: Изд-во ОмГТУ, 2004. Кн. 2. С. 255–258.
8. Борисовский П.А., Еремеев А.В. Генетический алгоритм для задачи о вершинном покрытии графа // Математика и информатика: наука и образование: межвуз. сб. науч. тр. Омск: Изд-во ОмГПУ, 2008. Вып. 7. С. 49–54.
9. Eremeev A.V. Hitting times of local and global optima in genetic algorithms with very high selection pressure // Yugoslav Journal of Operations Research. 2017. Vol. 27, Iss. 3. P. 323–339.
10. Reeves C., Rowe J. Genetic algorithms – principles and perspectives. A guide to GA theory. Kluwer Academic Publishers, 2003.
11. Blum C., Eremeev A., Zakharova Y. Hybridizations of evolutionary algorithms with Large Neighborhood Search // Computer Science Review. 2022. Vol. 46. Art. 100512.
12. Eremeev A.V., Kovalenko Y.V. A memetic algorithm with optimal recombination for the asymmetric travelling salesman problem // Memetic Computing. 2020. Vol. 12. P. 23–36.
13. Кочетов Ю.А., Младенович Н., Хансен П. Локальный поиск с чередующимися окрестностями // Дискретн. анализ и исслед. опер. 2003. Т. 10, № 1. С. 11–43.
14. Леванова Т.В., Федоренко А.С. Локальный поиск с чередующимися окрестностями для двухстадийной задачи размещения // Дискретн. анализ и исслед. опер. 2008. Т. 15, № 3. С. 43–57.
15. Levanova T.V., Gnusarev A.Yu. Variable neighborhood search algorithms for a competitive location problem with elastic demand // Diskretn. Anal. Issled. Oper. 2020. Vol. 27, No. 4. P. 80–103; J. Appl. Industr. Math. 2020. Vol. 14, No. 4. H. 693–705.
16. Eremeev A., Gette A., Hrushev S., Levanova T. Location and Design of Ground Stations for Software Defined Satellite Networks // Dynamics of Systems, Mechanisms and Machines (Dynamics). Omsk, Russian Federation, 2022. P. 1–4.

GENETIC ALGORITHMS FOR THE LOCATION OF GROUND STATIONS FOR A GLOBAL SATELLITE COMMUNICATIONS NETWORK

T.V. Levanova^{1,2}

Ph.D. (Phys.-Math.), Associate Professor, e-mail: levanovat@gmail.com

A.V. Gette²

Ph.D. Student, e-mail: gette4aleks@yandex.ru

¹Dostoevsky Omsk State University, Omsk, Russia

²Sobolev Institute of Mathematics of SB RAS, Omsk Division, Omsk, Russia

Abstract. Global satellite networks provide communication in various parts of the Earth. To ensure communication in large areas, it is necessary to locate ground stations that interact with satellite systems. The solution to the problem of location and design of ground stations can be divided into two stages. At the first stage, a set of ground stations under construction and the configuration of their equipment are determined. At the second stage, data packet routes are regularly calculated in each time interval, taking into account requests from customers and the current state of the communication network. The quality of customer service in a time interval is measured as the average number of lost or unrouted packets per unit of time. Previously, when solving the routing problem, it was assumed that all stations and transmitting devices

had already been selected. In this paper, the problem is considered as a two-stage one, in which it is necessary to solve the issues of the first and second stages together. The criterion is to minimize the total number of lost packets. To do this, a genetic algorithm and a genetic local search algorithm were built, their experimental study was performed, and a comparison with other algorithms was carried out.

Keywords: routing problem, location, genetic algorithm, local search, software defined network.

Дата поступления в редакцию: 18.02.2024

ОБЗОР МЕТОДОВ РЕГИСТРАЦИИ ИЗОБРАЖЕНИЙ НА ОСНОВЕ НЕЖЁСТКИХ ПРЕОБРАЗОВАНИЙ

С.В. Лейхтер

старший преподаватель, e-mail: leykhter@mail.ru

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

Аннотация. В статье приведён обзор способов регистрации изображений. Показано современное состояние проблемы. Предложены новые области применения алгоритмов регистрации изображений.

Ключевые слова: регистрация изображений, ключевые точки, характерные признаки, модели деформации, нежёсткие преобразования, диффеоморфизмы, топологический анализ данных.

Введение

Под регистрацией изображений понимается нахождение соответствия между искажёнными изображениями и эталонным изображением путём их приведения к единой системе координат. Искажения обусловлены получением изображений с разных устройств, в разное время, с разных точек обзора и т. п. Регистрация изображений является важной задачей компьютерного зрения и, как правило, выступает предварительным этапом алгоритмов: распознавания объектов, роботизированного составления карты/навигации, сшивания изображений, трёхмерного моделирования, трекинга, идентификации целей, медицинской визуализации и др.

1. Алгоритмы регистрации изображений

Для того чтобы изображения, полученные в результате различных измерений, можно было сравнить или проанализировать, используют алгоритмы регистрации (выравнивания) изображений. Их условно разделяют на два класса: 1) анализ на основе уровней насыщенности изображений; 2) выявление характерных признаков (ключевых точек) на изображениях [1]. Одно из изображений принимается подвижным – исходным (source), другие неподвижными – целевыми (target). Задача регистрации заключается в пространственном преобразовании исходного изображения для его выравнивания с целевым. Система отсчёта для целевого изображения принимается неподвижной [1]. Методы, основанные на «уровнях», находят корреляционную зависимость между графиками уровней (гистограммами) изображений или частей изображений (подизображений). Методы, основанные на «признаках», устанавливают соответствие между наборами ключевых точек на изображениях. Зная соответствие между наборами точек, определяется геометрическое преобразование

для сопоставления целевого изображения с эталонным [1]. Также известны комбинированные способы регистрации изображений, сочетающие возможности обоих методов [2].

2. Модели преобразований

Алгоритмы регистрации изображений различаются по типу преобразования пространства, которое они используют для совмещения изображений. Первый тип преобразований – широко используемые линейные преобразования, включающие перенос, масштабирование, поворот и другие аффинные преобразования. Линейные преобразования действуют на всё изображение и не могут учитывать локальные геометрические отличия между изображениями [1]. Второй тип преобразований допускает «эластичные» или «нежёсткие» преобразования. Эти преобразования подразумевают локальную деформацию целевого изображения для его совмещения с целевым. Нежёсткие преобразования включают радиальные базисные функции (TPS-сплайны, мультиквадратичные, с компактным носителем [1]), физические модели (механика сплошных сред), диффеоморфизмы и т. п. Преобразования обычно параметризуются, количество параметров определяется используемой моделью. Например, преобразование изображения может быть задано фиксированным числом параметров – вектором переноса, матрицей поворота и т. д. Такие модели называются параметрическими. Существуют модели, которые позволяют произвольно смещать каждый элемент изображения, так называемые непараметрические модели [3].

3. Определение характерных признаков

Таким образом, алгоритмы регистрации изображений работают с выбранными характерными признаками, определяемыми детекторами. Это могут быть: детекторы границ Кэнни, Дериша, Собеля; детекторы углов Харриса, Ши–Томаши, FAST; blob-детекторы: лапласиан гауссиана (LoG), разность гауссианов (DoG), определитель Гессе (DoH) [4], MSER [5]; аффинно-инвариантное выявление признаков [6, 7]; выявление локальных дескрипторов: SIFT, SURF, KAZE, гистограмма направленных градиентов (HOG), ORB и др. [8–11].

4. Методы пространственной и частотной области

Некоторые алгоритмы сопоставления объектов в пространственной области являются результатом развития традиционных способов регистрации изображений, в которых оператор выбирает контрольные точки на изображениях вручную. Когда выбрано достаточное количество точек для нахождения преобразования определённого типа, используются итеративные алгоритмы, такие как RANSAC [12], для надёжной (робастной) оценки параметров этого преобразования (например, аффинного). Методы частотной области работают для простых преобразований, таких как перенос, масштабирование, поворот. Применение метода фазовой корреляции к паре изображений позволяет определить относительное смещение между ними. При этом метод фазовой корреляции устойчив к шуму, перекрытиям и другим дефектам,

свойственным, например, медицинским или спутниковым снимкам. В основе метода лежит быстрое преобразование Фурье, которое не требует больших вычислительных затрат. Метод может быть расширен для определения различий в повороте и масштабировании между двумя изображениями [13, 14].

5. Одномодальные и мультимодальные методы

Также различают одномодальные и мультимодальные методы. Одномодальные методы подразумевают получение изображений одним и тем же типом датчика. Мультимодальные методы регистрируют изображения, полученные разными типами датчиков. Например, в медицинской визуализации часто объединяют результаты КТ, МРТ и ПЭТ изображений всего тела для локализации опухоли или результаты КТ с контрастным усилением и без него для сегментации анатомических образований [15].

6. Деформационные модели

Для нежёстких преобразований пространства могут использоваться геометрические преобразования, описываемые физическими моделями – моделями деформации. Выбор модели деформации имеет большое значение для процесса регистрации, поскольку от этого зависит как точность описания преобразования, так и вычислительная сложность процесса. Параметры, которые оценивает регистрация посредством стратегии оптимизации, соответствуют степеням свободы модели. Их количество варьируется от шести (в случае глобальных жёстких преобразований) до миллионов (в случае нежёстких преобразований) [3].

1. Модели деформации упругих твёрдых тел:

– в линейных моделях деформация задаётся уравнениями в частных производных Навье–Коши [16]:

$$\mu \nabla^2 u + (\mu + \lambda) \nabla(\nabla \cdot u) + F = 0,$$

где F – силовое поле, управляющее процессом регистрации на основе критерия соответствия изображений; μ – задаёт жёсткость материала (модуль сдвига); λ – первый коэффициент Ламе. Для вычисления перемещения узлов элементов используется метод конечных элементов. Главным ограничением линейных моделей упругости является их неспособность справляться со значительными деформациями;

– для описания больших деформаций были предложены нелинейные модели – модель гиперупругого материала Сен-Венана–Кирхгофа.

2. Модели движения жидкости: уравнения Навье–Стокса с очень низким числом Рейнольдса [17]:

$$\mu_f \nabla^2 v + (\mu_f + \lambda_f) \nabla(\nabla \cdot v) + F = 0,$$

первый член уравнения Навье–Стокса заставляет соседние точки деформироваться путём пространственного сглаживания поля скоростей. Поле скорости связано с полем перемещения следующим образом: $v(x;t) = \partial_t u(x;t) + (\nabla u(x;t)v(x;t))$. Поле скоростей интегрируется для оценки поля смещения.

Приведённые выше модели гарантируют сохранение топологии.

3. Модели на основе уравнения диффузии. Данный вид моделей использует тот факт, что гауссово ядро является функцией Грина уравнения диффузии (при соответствующих начальных и граничных условиях) [3]. Это позволяет обеспечить эффективную регуляризацию посредством свёрток с гауссовым ядром.

4. Модели, основанные на оценке кривизны. Известно, что методы нежёсткой регистрации изображений могут плохо сходиться, если начальное положение изображения недостаточно близко к решению. Как правило, такие методы подразумевают предварительный этап аффинного (жёсткого) выравнивания. Существуют методы, которые определяют поле деформации путём минимизации подходящей меры с учётом ограничения, основанного на кривизне сплайнов. Помимо точного и гладкого решения, автоматически выполняется жёсткое выравнивание. В основе схемы лежит численное решение уравнений Эйлера–Лагранжа [18].

5. Потоки диффеоморфизмов также могут задавать модель деформации [17–23]. В этом случае, деформация описывается эволюцией во времени уравнений Лагранжа. Гладкость поля скоростей ограничивается регуляризацией:

$$R = \int_0^1 \|v_t\|_V^2 dt.$$

При этом обеспечивается сохранение топологии и обратимость преобразования [17, 23]. Существует ряд программ, которые генерируют диффеоморфные преобразования координат на основе диффеоморфного отображения: Deformetrica [23], ANTS, DARTEL, DEMONS, LDDMM, StationaryLDDMM, MRICloud и др.

7. Топологический анализ данных

Ведутся теоретические исследования в новой области для задач анализа данных и компьютерного зрения, известной как топологический анализ данных (TDA). Одним из основных методов топологического анализа данных являются персистентные гомологии. Идея метода заключается в анализе топологических (симплициальных) комплексов инструментами алгебраической топологии. Симплициальный комплекс строится на основе облака точек [24]. Инструмент может применяться для экспериментальных данных различной природы, в том числе изображений [25]. Методы топологического анализа данных реализованы в библиотеках javaPlex, Gudhi [26], scikit-TDA, giotto-tda и др.

Заключение

Ввиду широкого спектра прикладных задач, требующих регистрацию изображений, не существует общего метода, оптимизированного для всех применений. Выбор или определение методики регистрации, меры сходства (взаимная корреляция, функция затрат) зависит от решаемой задачи. Поскольку множество решаемых задач компьютерного зрения постоянно расширяется, вовлекая в себя всё более разнообразные области применения, такие как, например, анализ цифровой модели рельефа поверхности полупроводниковых пластин [27], вопрос о выборе наилучшего

способа анализа изображений остаётся актуальным. К тому же появляются новые методы анализа данных, эффективность применения которых необходимо исследовать.

Литература

1. Goshtasby A.A. 2-D and 3-D image registration: for medical, remote sensing, and industrial applications. John Wiley & Sons, 2005.
2. Papademetris X., Jackowski A.P., Schultz R.T., Staib L.H., Duncan J.S. Integrated intensity and point-feature nonrigid registration // Medical Image Computing and Computer-Assisted Intervention—MICCAI 2004: 7th International Conference, Saint-Malo, France, September 26–29, 2004: Proceedings. Berlin; Heidelberg: Springer, 2004. P. 763–770.
3. Sotiras A., Davatzikos C., Paragios N. Deformable medical image registration: A survey // IEEE transactions on medical imaging. 2013. Vol. 32, No. 7. P. 1153–1190.
4. Lindeberg T. Scale-Space // Wiley Encyclopedia of Computer Science and Engineering. Vol. IV / Ed. B. Wah. John Wiley and Sons, 2008. P. 2495–2504.
5. Matas J., Chum O., Urban M., Pajdla T. Robust wide-baseline stereo from maximally stable extremal regions // Image and vision computing. 2004. Vol. 22, No. 10. P. 761–767.
6. Mikolajczyk K., Schmid C. Scale & affine invariant interest point detectors // International journal of computer vision. 2004. Vol. 60. P. 63–86.
7. Mikolajczyk K., Tuytelaars T., Schmid C., Zisserman A., Matas J., Schaffalitzky F., Kadir T., Van Gool L. A comparison of affine region detectors // International journal of computer vision. 2005. Vol. 65. P. 43–72.
8. Lindeberg T. Edge detection and ridge detection with automatic scale selection // International journal of computer vision. 1998. Vol. 30. P. 117–156.
9. Lindeberg T. Image matching using generalized scale-space interest points // Journal of mathematical Imaging and Vision. 2015. Vol. 52. P. 3–36.
10. Lindeberg T. Scale invariant feature transform // Scholarpedia. 2012. Vol. 7 (5). Art. 10491.
11. Rublee E. et al. ORB: An efficient alternative to SIFT or SURF // 2011 International conference on computer vision. IEEE, 2011. P. 2564–2571.
12. Fischler M.A., Bolles R.C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography // Communications of the ACM. 1981. Vol. 24, No. 6. P. 381–395.
13. Reddy B.S., Chatterji B.N. An FFT-based technique for translation, rotation, and scale-invariant image registration // IEEE transactions on image processing. 1996. Vol. 5, No. 8. P. 1266–1271.
14. Zokai S., Wolberg G. Image registration using log-polar mappings for recovery of large-scale similarity and projective transformations // IEEE transactions on image processing. 2005. Vol. 14, No. 10. P. 1422–1434.
15. Ristea N.C., Miron A.-I., Savencu O., Georgescu M.-I., Verga N., Khan F.S., Ionescu R.T. CyTran: A cycle-consistent transformer with multi-level consistency for non-contrast to contrast CT translation // Neurocomputing. 2023. Vol. 538. Art. 126211.
16. Davatzikos C. Spatial transformation and registration of brain images using elastically deformable models // Computer Vision and Image Understanding. 1997. Vol. 66, No. 2. P. 207–222.

17. Christensen G.E., Rabbitt R.D., Miller M.I. Deformable templates using large deformation kinematics // IEEE transactions on image processing. 1996. Vol. 5, No. 10. P. 1435–1447.
18. Fischer B., Modersitzki J. Curvature based image registration // Journal of Mathematical Imaging and Vision. 2003. Vol. 18. P. 81–85.
19. Vaillant M., Glaunes J. Surface matching via currents // Biennial international conference on information processing in medical imaging. Berlin; Heidelberg: Springer, 2005. P. 381–392.
20. Miller M.I., Trounev A., Younes L. Geodesic shooting for computational anatomy // Journal of mathematical imaging and vision. 2006. Vol. 24. P. 209–228.
21. Younes L. Shapes and diffeomorphisms. Berlin: Springer, 2010.
22. Durrleman S., Prastawa M., Charon N., Korenberg J.R., Joshi S., Gerig G., Trounev A. Morphometry of anatomical shape complexes with dense deformations and sparse parameters // NeuroImage. 2014. Vol. 101. P. 35–49.
23. Bône A., Louis M., Martin B., Durrleman S. Deformetrica 4: an open-source software for statistical shape analysis // Shape in Medical Imaging: International Workshop, ShapeMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018: Proceedings. Springer International Publishing, 2018. P. 3–13.
24. Carlsson G. Topology and data // Bulletin of the American Mathematical Society. 2009. Vol. 46, No. 2. P. 255–308.
25. Чуканов С.Н., Чуканов И.С., Лейхтер С.В. Использование персистентной энтропии для топологического анализа данных // Математические структуры и моделирование. 2023. № 3 (67). С. 94–103.
26. Maria C., Boissonnat J.-D., Glisse M., Yvinec M. The gudhi library: Simplicial complexes and persistent homology // Mathematical Software–ICMS 2014: 4th International Congress, Seoul, South Korea, August 5-9, 2014: Proceedings. Berlin; Heidelberg: Springer, 2014. P. 167–174.
27. Дедкова А.А., Флоринский И.В., Дюжев Н.А. Методика исследования изменения формы пластин и тонкопленочных мембран с использованием геоморфометрических подходов // Журнал технической физики. 2022. Т. 92, № 8. С. 1113–1123.

OVERVIEW OF IMAGE REGISTRATION METHODS BASED ON NON-RIGID TRANSFORMATIONS

S.V. Leykhter

Assistant Professor, e-mail: leykhter@mail.ru

Dostoevsky Omsk State University, Omsk, Russia

Abstract. The article gives an overview of image registration methods. The modern state of the problem is shown. New fields of using the image registration algorithms are proposed.

Keywords: image registration, points of interest, local features, deformable model, non-rigid transformations, diffeomorphisms, TDA.

Дата поступления в редакцию: 27.02.2024

ИНЖИНИРИНГ ПРИВИЛЕГИЙ В ЗАДАЧЕ ПОСТРОЕНИЯ РОЛЕВОЙ ПОЛИТИКИ РАЗГРАНИЧЕНИЯ ДОСТУПА

Н.Ф. Богаченко

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

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

Аннотация. Проблема инжиниринга ролей расширяется подзадачей разработки привилегий. Предполагается, что в информационной системе задано дискреционное разграничение доступа. Для построения ролевой политики безопасности предлагается методика, основанная на алгоритмах анализа формальных понятий. По матрице доступов строится решётка Галуа, узлы которой интерпретируются как возможные привилегии. Определяются критерии выбора оптимального набора привилегий и обсуждается эвристический алгоритм решения поставленной задачи.

Ключевые слова: разграничение доступа, роли, привилегии, матрица доступов, анализ формальных понятий.

Одним из основных способов обеспечения информационной безопасности является защита от несанкционированного доступа. При этом задача построения политики разграничения доступа считается не менее важной, чем, например, стойкость используемых криптографических алгоритмов. Практическая реализация классических моделей управления доступом сталкивается с рядом трудностей применительно к крупномасштабным информационным системам (Large-Scale Complex IT Systems). В данной работе будет рассмотрена ролевая модель разграничения доступа и проблемы, возникающие при её построении.

1. Разработка ролей

Рассмотрим математическую постановку задачи построения ролевого разграничения доступа (Role-Based Access Control) [1] в некоторой информационной системе. Этот процесс состоит из трёх этапов:

1. Построение иерархии ролей:

- ✓ формирование множества ролей: $R = \{r_1, \dots, r_n\}$;
- ✓ формирование множества привилегий: $P = \{p_1, \dots, p_m\}$;
- ✓ назначение привилегий ролям: $RP : R \rightarrow 2^P$;
- ✓ авторизация ролей: $RR : R \rightarrow 2^R$.

2. Предоставление ролей пользователям:

- ✓ формирование множества пользователей: $U = \{u_1, \dots, u_s\}$;
- ✓ назначение привилегий пользователям: $UP : U \rightarrow 2^P$;
- ✓ авторизация пользователей: $UR : U \rightarrow 2^R$.

3. Работа в системе:

- ✓ формирование множества сеансов работы: $C = \{c_1, \dots, c_t\}$;
- ✓ управление доступом: $CU : C \times U \rightarrow 2^R$.

При реализации ролевой политики разграничения доступа в небольшой информационной системе, как правило, не возникает сложности в выборе множеств R и P и отображения RP . Другими словами, определение ролей и закрепление за каждой ролью набора привилегий осуществляется администратором безопасности в ручном режиме на основе анализа бизнес-процессов. Ситуация резко усложняется, когда информационная система является крупномасштабной [2, 3]. В этом случае поиск оптимального (корректного, полного) множества ролей и связанных с каждой ролью привилегий является весьма сложной задачей, требующей вычислительной поддержки (автоматизации). Данная задача получила название «инжиниринг ролей» (Role Engineering) [4]. Как известно, для решения сложных проблем, в том числе и для задачи инжиниринга ролей, возможны два подхода, два пути решения: нисходящий (top-down) и восходящий (bottom-up).

Применительно к задаче инжиниринга ролей нисходящий подход (или, как его ещё называют, метод «с чистого листа») заключается в классической последовательности поэтапного построения ролевой политики разграничения доступа. Группа экспертов определяет множество R , задаёт отображения RP и RR . И лишь затем происходит переход ко второму этапу – к авторизации пользователей на роли. Но в крупномасштабных информационных системах как число пользователей, так и число объектов доступа (а значит, и число привилегий) существенно увеличивается. Кроме того, нередко необходим учёт уже имеющихся информационных потоков, разрешений и запретов на доступ и т. п. В связи с этим всё более востребован восходящий подход к решению задачи инжиниринга ролей. В этом случае построение ролевой модели начинается со второго этапа: за основу берутся множество U и отображение UP , т. е. уже существующие правила доступа – потребности пользователей в определённом наборе привилегий. При таком подходе к проектированию ролевого разграничения доступа часто используются методы интеллектуального анализа данных (Data Mining). В связи с чем задачу инжиниринга ролей, для решения которой используется восходящий подход, принято называть «проблемой разработки ролей» (Role Mining Problem) [5]. Постановка этой задачи может быть формализована следующим образом [6]:

- Дано:
 - ✓ множество пользователей: $U = \{u_1, \dots, u_s\}$;
 - ✓ множество привилегий: $P = \{p_1, \dots, p_m\}$;
 - ✓ отображение, определяющее назначение привилегий пользователям: $UP : U \rightarrow 2^P$.
- Найти:
 - ✓ множество ролей: $R = \{r_1, \dots, r_n\}$;
 - ✓ отображение, определяющее назначение привилегий ролям: $RP : R \rightarrow 2^P$;
 - ✓ отображение, определяющее авторизацию пользователей: $UR : U \rightarrow 2^R$;
 - ✓ отображение, определяющее авторизацию ролей: $RR : R \rightarrow 2^R$.

Следует заметить, что в рассмотренной, уже ставшей классической, постановке проблемы разработки ролей в числе исходных данных определены привилегии, необходимые каждому пользователю для работы в системе, т. е. весь набор привилегий P и их желаемое распределение между пользователями UP известны заранее.

Возникает вопрос, что делать, если в крупномасштабной информационной системе множество привилегий P и тем более отображение UP ещё не определены? Если подсистема разграничения доступа строится «с нуля», то здесь необходима серьёзная работа экспертов. Но в случае, когда в информационной системе ранее уже была реализована дискреционная политика разграничения доступа (Discretionary Access Control) [7], являющаяся базовой для всех защищённых компьютерных систем, представляется возможным частично автоматизировать процесс выявления множества привилегий P за счёт применения методов интеллектуального анализа данных (используемых для решения проблемы разработки ролей в целом).

2. Вспомогательные модели и методы

Дискреционное разграничение доступа основано на произвольном управлении доступом: разрешение на доступ определяется для каждого субъекта к каждому объекту. Пусть S – множество субъектов, O – множество объектов, A – множество видов доступа. Для каждой пары $(s_i, o_j) \in S \times O$ задаётся набор разрешённых видов доступа $\alpha_{ij} \subseteq A$, т. е. определяется правило доступа (s_i, o_j, α_{ij}) . Такие правила организовываются в матрицу доступов \mathbf{M} размерности $|S| \times |O|$, в которой каждому субъекту соответствует своя строка, каждому объекту – свой столбец, а на их пересечении указывается набор разрешённых видов доступа α_{ij} .

Не ограничивая общности, будем считать, что $U = S$, т. е. множества пользователей и субъектов совпадают. В реальных системах $U \subseteq S$, но для решения наших задач достаточно преобразовать матрицу доступов \mathbf{M} , удалив из неё строки, которые не соответствуют пользователям.

Далее рассмотрим один из подходов к решению вопроса формирования множества привилегий P на основе имеющейся матрицы доступов \mathbf{M} . Предлагается использовать методы анализа формальных понятий (Formal Concept Analysis) [8], успешно применяемые для решения классической проблемы разработки ролей [6].

В основе анализа формальных понятий лежит определение формального контекста – тройки (I, J, \mathbf{R}) , где I – множество предметов; J – множество признаков; \mathbf{R} – бинарная матрица размерности $|I| \times |J|$, сопоставляющая признаки предметам. На любые подмножества $X \subseteq I$, $Y \subseteq J$ можно подействовать операторами Галуа:

$$X^\uparrow = \{j \in J | \forall i \in X : [\mathbf{R}]_{ij} = 1\},$$

$$Y^\downarrow = \{i \in I | \forall j \in Y : [\mathbf{R}]_{ij} = 1\}.$$

Несложно заметить, что X^\uparrow – это множество признаков, общих для предметов из X , Y^\downarrow – множество предметов, гарантированно имеющих признаки из Y . Формальное понятие формального контекста – это пара (X, Y) такая, что

$$(X = Y^\downarrow) \wedge (Y = X^\uparrow).$$

В формальном понятии (X, Y) подмножество X принято называть объёмом, а подмножество Y – содержанием. Очевидно, что каждое формальное понятие в матрице \mathbf{R} определяет заполненный единицами прямоугольник с точностью до перестановки строк и/или столбцов. Для любых двух формальных понятий (X_1, Y_1) и (X_2, Y_2) выполняется правило включения:

$$X_1 \subseteq X_2 \iff Y_1 \supseteq Y_2.$$

Это позволяет на множестве всех формальных понятий заданного формального контекста ввести отношение частичного порядка

$$(X_1, Y_1) \prec (X_2, Y_2) \iff X_1 \subseteq X_2 (Y_1 \supseteq Y_2)$$

и доказать, что формальные понятия, упорядоченные согласно этому отношению, образуют математическую решётку. Эта решётка называется решёткой Галуа и визуализируется в виде диаграммы Хассе (ориентированного графа), узлами которой являются формальные понятия. Чем выше в диаграмме находится формальное понятие (X, Y) , тем больше его объём X и меньше содержание Y .

3. Разработка привилегий

Исходя из подзадач, возникающих на практике при построении ролевой политики разграничения доступа в крупномасштабных информационных системах, очевидна необходимость в формализации и поиске способов решения проблемы выявления полного множества привилегий и закрепления этих привилегий за пользователями – «проблемы инжиниринга привилегий»:

- Дано: множество пользователей U .
- Найти:
 - ✓ множество привилегий P ;
 - ✓ отображение, определяющее назначение привилегий пользователям, $UP : U \rightarrow 2^P$.

Если для решения поставленной задачи применяются восходящие принципы проектирования, т. е. за основу решения берутся уже существующие в системе предпосылки к разграничению доступа и используются методы интеллектуального анализа данных, то задачу инжиниринга привилегий назовём «проблемой разработки привилегий» [9, 10]. Одна из возможных постановок проблемы разработки привилегий приведена далее.

- Дано:
 - ✓ множество пользователей U ;
 - ✓ множество объектов, к которым осуществляется доступ, O ;
 - ✓ множество видов доступа A ;
 - ✓ матрица доступов \mathbf{M} размерности $|U| \times |O|$.
- Найти:
 - ✓ множество привилегий P ;
 - ✓ отображение, определяющее назначение привилегий пользователям, $UP : U \rightarrow 2^P$.

Таким образом в процессе формирования (или выявления) элементов множества привилегий P за основу берутся разрешённые пользователям виды доступа к объектам системы. Важно напомнить, что привилегии в ролевой политике разграничения доступа выдаются пользователям на действия в информационной системе в целом, применительно к целой группе или классу объектов. Виды доступа в дискреционной политике, напротив, задают разрешения на доступ к определённому объекту. Поэтому можно сформулировать следующие требования к процедуре определения привилегий:

1. Привилегия должна быть соотнесена с группой объектов, к которым задан один и тот же вид доступа.
2. Привилегия должна быть выдана группе пользователей, которые имеют один и тот же вид доступа к некоторому объекту.

Эти правила приводят к идее использования методов анализа формальных понятий, которые позволят по матричному представлению исходных данных выделить привилегии с оптимальными наборами объектов и пользователей. Предлагается следующая методика решения проблемы разработки привилегий, состоящая из трёх этапов.

Этап 1. В соответствии с видами доступа, определёнными множеством A , построим бинарные матрицы $\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_{|A|}$ по правилу:

$$[\mathbf{M}_k]_{ij} = 1 \iff a_k \in [\mathbf{M}]_{ij},$$

где $k = 1, \dots, |A|$, $a_k \in A$, \mathbf{M} – матрица доступов. Данный процесс назовём декомпозицией матрицы \mathbf{M} по видам доступа. Далее матрицы, полученные в результате декомпозиции, будем рассматривать отдельно.

Этап 2. Следующая подзадача – найти оптимальный набор привилегий P_k для вида доступа a_k . Под оптимальным будем понимать такой набор, который в первую очередь минимизирует число выделенных привилегий. Но если руководствоваться только этим критерием, то достаточно будет создать одну привилегию, которая выдаст доступ a_k ко всем объектам множества O . К сожалению, при таком подходе подавляющее большинство пользователей получают доступ к объектам, к которым в матрице \mathbf{M}_k доступ был запрещён. С точки зрения минимизации рисков утечки информации естественно потребовать, чтобы число таких «лишних» объектов стремилось к нулю. А это приведёт к обратному эффекту – число привилегий придётся увеличивать. Наконец, выделенные привилегии должны обеспечить все доступы, которые разрешались матрицей \mathbf{M}_k .

Представим нашу проблему в терминах анализа формальных понятий. Для каждой бинарной матрицы \mathbf{M}_k определим формальный контекст $M_k = (U, O, \mathbf{M}_k)$, в котором множество предметов – это множество пользователей U , множество признаков – это множество объектов O , бинарная матрица, сопоставляющая предметы и признаки, – это матрица \mathbf{M}_k .

Далее найдём все формальные понятия (X_k^i, Y_k^i) формального контекста M_k и построим решётку Галуа Γ_k . Существуют различные алгоритмы решения этой задачи. Их классификацию и описание можно найти, например, в работе [11]. Узлы полученной решётки (формальные понятия (X_k^i, Y_k^i)) будем интерпретировать как возможные привилегии, выделенные для вида доступа a_k . Среди узлов решётки Γ_k

необходимо выбрать набор привилегий $P_k = \{(X_k^1, Y_k^1), \dots, (X_k^{m_k}, Y_k^{m_k})\}$, удовлетворяющий следующим критериям и ограничениям:

1. Число выделенных привилегий минимально:

$$F = |P_k| \longrightarrow \min. \quad (1)$$

2. Число «лишних» объектов, образующихся при переходе от матрицы доступов к распределению привилегий между пользователями, минимально:

$$G = \sum_{u_i \in U} \mathcal{L}(\mathbf{M}, P_k, UP) \longrightarrow \min, \quad (2)$$

здесь \mathcal{L} – оператор, вычисляющий количество «лишних» объектов¹.

3. Объединение содержаний выделенных привилегий покрывает множество объектов:

$$\bigcup_{(X_k^i, Y_k^i) \in P_k} Y_k^i = O. \quad (3)$$

Очевидно, что поиск точного решения сформулированной подзадачи выбора оптимального набора привилегий (1) – (3) весьма трудоёмкий. Он сводится к полному перебору возможных подмножеств привилегий, удовлетворяющих требованиям матрицы доступов \mathbf{M}_k в разрезе разрешений на доступ (требованиям ограничения (3)). С практической точки зрения достаточно найти приближённое решение. Рассмотрим эвристический алгоритм выбора набора P_k по матрице \mathbf{M}_k и по решётке Γ_k .

Несложно понять, что для каждой j -й строки матрицы \mathbf{M}_k , отвечающей пользователю u_j , найдётся узел (X_k^i, Y_k^i) в решётке Γ_k , в котором:

- объём $X_k^i = \{u_j\}$ состоит из одного пользователя;
- содержание Y_k^i включает те объекты, которым отвечают единичные клетки выбранной строки матрицы \mathbf{M}_k .

На первом шаге отбора привилегий необходимо найти минимальное число именно таких узлов решётки Γ_k , объёмы которых содержат по одному пользователю, а все содержания в объединении покрывают множество объектов O . Тем самым будут удовлетворены все разрешения на доступ, выдаваемые матрицей \mathbf{M}_k (ограничение (3)). Но этот процесс эквивалентен поиску минимального числа строк матрицы \mathbf{M}_k , покрывающих единицами все её столбцы. Эта подзадача представляет собой интерпретацию хорошо известной задачи минимизации булевой формулы методом Квайна–мак-Класки в пункте работы с импликантной матрицей, алгоритмы решения которой изучены и реализованы [12]. От выбранных строк опять перейдём к узлам решётки Γ_k .

На втором шаге, с целью минимизации числа «лишних» объектов при выдаче привилегий пользователям, для каждого выбранного на предыдущем шаге узла (X_k^i, Y_k^i) рассмотрим узлы, из которых ведут в него дуги, т. е. рассмотрим «родителей» этого узла. Если среди узлов-«родителей» можно выбрать набор узлов, содержания которых совместно покрывают содержание узла (X_k^i, Y_k^i) , то фиксируем

¹Вопросы построения отображения UP и оператора \mathcal{L} по матрице доступов \mathbf{M} и набору привилегий P_k требуют отдельного обсуждения и выходят за рамки данной статьи.

такой минимальный по числу узлов набор и привилегию (X_k^i, Y_k^i) заменяем на привилегии, соответствующие отобранным узлам. Этот процесс, назовём его расщеплением привилегий, ведёт к увеличению числа привилегий и в зависимости от структуры решётки Γ_k может итеративно продолжаться вплоть до её верхней грани. Критерием остановки следует выбрать условие достижения некоторого баланса между числом привилегий и числом «лишних» объектов при авторизации пользователей на привилегии.

Этап 3. Пусть для каждого вида доступа $a_k \in A$ определён набор привилегий P_k . Сформируем множество привилегий $P = \bigcup P_k$ ($k = 1, \dots, |A|$) и определим отображение $UP : U \rightarrow 2^P$ так, чтобы обеспечить все разрешённые матрицей \mathbf{M} доступы².

Пример 1. Разберём второй этап решения проблемы разработки привилегий на примере. Пусть построен формальный контекст M_k с матрицей \mathbf{M}_k .

$$\mathbf{M}_k = \begin{matrix} & o_1 & o_2 & o_3 & o_4 & o_5 & o_6 & o_7 \\ \begin{matrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}.$$

Иерархия возможных привилегий, порождённая решёткой формальных понятий Γ_k формального контекста M_k , представлена на рис. 1. Возможные привилегии имеют следующие объёмы и содержания:

- $p_1 = (\{u_1, u_2, u_3, u_4, u_5\}, \emptyset)$; $p_2 = (\{u_1, u_2, u_3\}, \{o_1, o_6\})$;
- $p_3 = (\{u_2, u_4\}, \{o_3, o_5\})$; $p_4 = (\{u_4, u_5\}, \{o_2\})$;
- $p_5 = (\{u_2, u_3\}, \{o_1, o_4, o_6\})$; $p_6 = (\{u_2\}, \{o_1, o_3, o_4, o_5, o_6\})$;
- $p_7 = (\{u_4\}, \{o_2, o_3, o_5\})$; $p_8 = (\{u_5\}, \{o_2, o_7\})$;
- $p_9 = (\emptyset, \{o_1, o_2, o_3, o_4, o_5, o_6, o_7\})$.

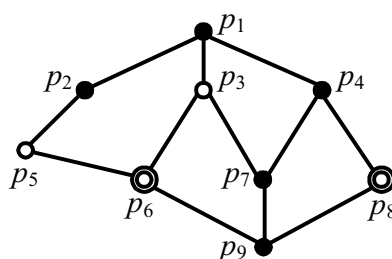


Рис. 1. Диаграмма Хассе решётки Галуа Γ_k

На первом шаге отбора привилегий будут выбраны привилегии p_6 и p_8 , так как их объёмы состоят из одного пользователя, а объединение содержаний покрывает всё множество объектов. На втором шаге обнаруживается возможность заменить привилегию p_6 её «родителями» – привилегиями p_3 и p_5 , так как содержания этих двух

²Более подробно алгоритм построения отображения UP будет рассмотрен в последующих публикациях.

привилегий в объединении покрывают содержание привилегии p_6 . Таким образом, имеются два варианта формирования набора P_k : $P_k^1 = \{p_6, p_8\}$ и $P_k^2 = \{p_3, p_5, p_8\}$. Оценим число выделенных привилегий и число «лишних» объектов для каждого варианта (см. табл. 1, 2). Для P_k^1 : $F = 2, G = 9$; для P_k^2 : $F = 3, G = 2$. Окончательный выбор, какое из подмножеств P_k^1 или P_k^2 принять за искомый набор P_k , осуществляется экспертами, исходя из того, какой критерий – F или G – более значим.

Таблица 1. Отображение UP и «лишние» объекты для набора привилегий P_k^1

Пользователь	Привилегии	«Лишние» объекты
u_1	p_6	$\{o_3, o_4, o_5\}$
u_2	p_6	\emptyset
u_3	p_6	$\{o_3, o_5\}$
u_4	$p_6 \cup p_8$	$\{o_1, o_4, o_6, o_7\}$
u_5	p_8	\emptyset

Таблица 2. Отображение UP и «лишние» объекты для набора привилегий P_k^2

Пользователь	Привилегии	«Лишние» объекты
u_1	p_5	$\{o_4\}$
u_2	$p_3 \cup p_5$	\emptyset
u_3	p_5	\emptyset
u_4	$p_3 \cup p_8$	$\{o_7\}$
u_5	p_8	\emptyset

Заключение

Результаты, представленные в статье, носят теоретический характер и затрагивают математическую постановку обсуждаемых проблем, алгоритмическое описание возможных путей решения. Практическая реализация предлагаемых алгоритмов и анализ результатов вычислительного эксперимента – это задачи дальнейших исследований.

Литература

1. 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.
2. Богаченко Н.Ф. Анализ проблем управления разграничением доступа в крупномасштабных информационных системах // Математические структуры и моделирование. 2018. № 2 (46). С. 135–152.

3. Богаченко Н.Ф. О сложности подсистем разграничения доступа крупномасштабных информационных систем // Математические структуры и моделирование. 2018. № 4 (48). С. 92–98.
4. Coyne E.J. Role engineering // RBAC '95: Proceedings of the first ACM Workshop on Role-based access control. New York: ACM Press, 1995. P. 4–5.
5. Kuhlmann M., Shohat D., Schimpf G. Role mining – revealing business roles for security administration using data mining technology // SACMAT '03: Proceedings of the eighth ACM symposium on Access control models and technologies. 2003. P. 179–186.
6. 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.
7. Harrison M.A., Ruzzo W.L., Ullman J.D. On Protection in Operating Systems // Communications of the ACM. 1975. P. 14–25.
8. Wille R. Restructuring Lattice Theory: an approach based on hierarchies of concept // Ordered sets / Ed. I. Rival. Dordrecht; Boston: Reidel, 1982.
9. Богаченко Н.Ф. Интеллектуальный анализ политик разграничения доступа больших информационных систем // Математическое и компьютерное моделирование: сб. материалов V Междунар. науч. конф. Омск: Изд-во Ом. гос. ун-та, 2017. С. 142–145.
10. Богаченко Н.Ф. Проблема разработки полномочий // Омские научные чтения – 2020: материалы Четвёртой Всерос. науч. конф. Омск: Изд-во Ом. гос. ун-та, 2020. С. 367–369.
11. Семенова Д.В., Катаева А.В., Монгуш Ч.М. Метод декомпозиции формального контекста и неизбыточное представление закономерностей в многомерных данных: моногр. Красноярск: Сиб. федер. ун-т, 2020.
12. Дмитриев Г.А., Комиссарчик В.Ф., Марголис Б.И. Программа минимизации функций алгебры логики методом Мак-Класки // Программные продукты и системы. 1997. № 2. URL: <https://swsys.ru/index.php?page=article&id=1032> (дата обращения: 21.02.2024).

PERMISSIONS ENGINEERING IN THE TASK OF CONSTRUCTING A ROLE-BASED ACCESS CONTROL POLICY

N.F. Bogachenko

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

Dostoevsky Omsk State University, Omsk, Russia

Abstract. The problem of role engineering is extended by the subproblem of permissions engineering. It is assumed that discretionary access control is specified in the information system. To build a role-based security policy, a technique based on algorithms for formal concept analysis is proposed. Based on the access matrix, a Galois lattice is constructed, the nodes of which are interpreted as possible permissions. Criteria for choosing the optimal set of permissions are determined and a heuristic algorithm for solving the problem is discussed.

Keywords: access control, roles, permissions, access matrix, formal concept analysis.

Дата поступления в редакцию: 26.02.2024

ОПТИМИЗАЦИЯ ВЫБОРА СТРАТЕГИИ ЗАЩИТЫ ОТ ВРЕДОНОСНЫХ АТАК НА ОСНОВЕ РЕШЕНИЯ БИМАТРИЧНОЙ ИГРЫ С УЧЁТОМ РИСКОВ

Т.В. Вахний

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

С.В. Вахний

студент, e-mail: vakhniysv@mail.ru

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

Аннотация. В статье описано решение биматричной игры между администратором безопасности и атакующим злоумышленником в два этапа. На первом этапе предлагается применять критерии оптимальности, учитывающие цену игры, а на втором этапе – критерии, учитывающие риски игроков. Такой подход помогает находить наилучшую стратегию защиты от атак злоумышленника среди стратегий с одинаковой ценой игры.

Ключевые слова: компьютерная система, цифровизация, кибербезопасность, биматричная игра, оптимальная стратегия.

Введение

В современном мире цифровая трансформация охватывает почти все сферы деятельности, от бизнеса и образования до здравоохранения и государственного управления. Новейшие технологии, такие как искусственный интеллект, большие данные и интернет вещей, открывают огромные возможности для создания новых продуктов и услуг. Благодаря этому глобальная цифровизация позволяет компаниям стать более эффективными, инновационными и адаптивными, что является ключевым фактором успешной конкуренции на рынке. В связи с этим вопросы кибербезопасности и сохранности данных к кибератакам выходят на первый план. Непрерывно растущий рост числа атак, их усложнение и изощрённость побуждают к созданию большого количества средств защиты, и построение надёжной системы безопасности становится всё более сложной задачей.

Стохастической природе проблем защиты компьютерных систем соответствуют математические методы принятия решений в условиях неопределённости, в частности методы теории игр. Использование теоретико-игрового подхода позволяет анализировать взаимодействие между администратором безопасности и атакующим злоумышленником, на основе чего обеспечить оптимизацию выбора программных продуктов для построения наилучшей системы безопасности компании [1–3].

В данной статье для оптимизации защиты компьютерной системы компании предлагается построить биматричную игру администратора безопасности со злоумышленником и предусмотреть нахождение её решения в два этапа. При этом на

первом этапе в расчётах следует применять критерии оптимальности, традиционно учитывающие цену игры, а на втором этапе – критерии, учитывающие риски игроков. Такой подход может быть полезен для нахождения наилучшей стратегии защиты от атак злоумышленника среди стратегий с одинаковой ценой игры.

1. Постановка задачи и игровой подход

Один из подходов, моделирующих игру администратора безопасности и атакующего злоумышленника, основан на проведении биматричной игры, в которой интересы игроков не совпадают и задаются разными платёжными матрицами. В платёжных матрицах строки соответствуют стратегиям одного игрока (программное средство или набор из программных средств), а столбцы – стратегиям другого игрока, на их пересечении в первой платёжной матрице A стоит цена игры для администратора безопасности, а во второй платёжной матрице B – цена игры для злоумышленника. Проведение биматричной игры позволяет определить наиболее выигрышные стратегии для каждого игрока.

Если администратор для обеспечения безопасности компьютерной системы компании может выбирать из S программных средств защиты, и при этом их можно устанавливать одновременно, то у него будет $N = (2^S - 1)$ вариантов стратегий. Аналогично, если злоумышленник имеет L способов атаки, то у него будет $M = (2^L - 1)$ вариантов стратегий.

Таблица 1. Платёжная матрица A

	y_1	y_2	...	y_M
x_1	a_{11}	a_{12}	...	a_{1M}
x_2	a_{21}	a_{22}	...	a_{2M}
...
x_N	a_{N1}	a_{N2}	...	a_{NM}

Таблица 2. Платёжная матрица B

	y_1	y_2	...	y_M
x_1	b_{11}	b_{12}	...	b_{1M}
x_2	b_{21}	b_{22}	...	b_{2M}
...
x_N	b_{N1}	b_{N2}	...	b_{NM}

Ходом администратора безопасности является использование одной из N стратегий защиты x_i ($i = 1, 2, \dots, N$), а ходом злоумышленника – применение одной из M стратегий атаки y_j ($j = 1, 2, \dots, M$) на компьютерную систему компании. Последовательно перебирая все стратегии игроков, можно заполнить две таблицы, в одной из них указывая ущерб администратора a_{ij} (см. табл. 1), а во второй – прибыль b_{ij} злоумышленника (см. табл. 2), соответственно, при выборе стратегии защиты x_i и способа атаки y_j . Из табл. 1 и 2 можно выписать платёжные матрицы A и B , содержащие N строк и M столбцов с элементами a_{ij} и b_{ij} соответственно:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1M} \\ a_{21} & a_{22} & \dots & a_{2M} \\ \dots & \dots & \dots & \dots \\ a_{N1} & a_{N2} & \dots & a_{NM} \end{pmatrix}; \quad B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1M} \\ b_{21} & b_{22} & \dots & b_{2M} \\ \dots & \dots & \dots & \dots \\ b_{N1} & b_{N2} & \dots & b_{NM} \end{pmatrix}.$$

Здесь элементы a_{ij} платёжной матрицы администратора безопасности вычисляются следующим образом:

$$a_{ij} = R(x_i, y_j) + G_i,$$

где $R(x_i, y_j)$ – величина ущерба от атаки y_j при использовании стратегии защиты x_i ; G_i – затраты администратора на приобретение и использование программных средств защиты, необходимых для реализации стратегии x_i .

Аналогично элементы b_{ij} платёжной матрицы злоумышленника вычисляются по формуле:

$$b_{ij} = P(x_i, y_j) - F_j,$$

где $P(x_i, y_j)$ – величина прибыли от вредоносной атаки y_j при использовании администратором стратегии защиты x_i ; F_j – затраты злоумышленника на реализацию атаки y_j .

Биматричная игра состоит в том, что администратор выбирает стратегию защиты x_i , злоумышленник выбирает стратегию атаки y_j , после чего вычисляется исход игры, заключающийся в том, что администратор терпит ущерб, равный a_{ij} , а злоумышленник получает прибыль b_{ij} . Цель администратора безопасности – выбор такой стратегии, т. е. набора программных средств защиты, который сводит потери от атак и затраты на покупку средств защиты к минимуму, а цель атакующего – выбор такой стратегии, которая даст ему наибольший выигрыш. Решение биматричной игры сводится к отысканию равновесных (оптимальных) стратегий игроков x_{i_0} и y_{j_0} . Выбор одним из игроков любой другой стратегии вероятнее всего приведёт к ухудшению его результатов игры и улучшению их у противника.

В учебной литературе биматричная игра описывается как одноходовая [1,4]. Однако при выборе из большого количества программных продуктов у нескольких найденных наилучших стратегий администратора безопасности цена игры может оказаться одинаковой. Для таких случаев предлагается предусмотреть второй этап для уточнения решения биматричной игры. На первом этапе нужно найти наилучшие стратегии игроков и проверить, есть ли среди них такие, которые мало отличаются по цене игры. Если таковые найдутся, то предлагается построить новые платёжные матрицы из найденных на первом этапе наилучших стратегий игроков и уточнить решение биматричной игры, используя в этот раз критерии оптимальности, основанные на оценке степени удачности применений стратегий, т. е. величины риска. Риск рассчитывается как разность между ценой игры при выборе стратегии в условиях, когда заранее не известна стратегия второго игрока, и в условиях, когда она заранее известна [4].

2. Построение матриц рисков для администратора безопасности и злоумышленника

Матрица рисков для администратора безопасности R строится по столбцам его платёжной матрицы A . В каждом столбце нужно найти наименьшее значение, это значение по очереди вычестить из всех значений в данном столбце и результат записать в те же позиции. Элементы матрицы рисков для администратора безопасности рассчитывают по формуле: $r_{ij} = a_{ij} - \min_j a_{ij}$. Они показывают, насколько больше

может быть ущерб компьютерной системе по сравнению с минимально возможным значением для каждого типа атаки злоумышленника из-за неверного выбора стратегии защиты.

Аналогично матрица рисков для атакующего злоумышленника строится по строкам его платёжной матрицы B . В каждой строке нужно найти наибольшее значение, из этого значения по очереди вычесть все значения в данной строке и результат записать в те же позиции. Элементы матрицы рисков для злоумышленника рассчитывают по формуле: $v_{ij} = \max_i b_{ij} - b_{ij}$. Они показывают, насколько меньше может быть выигрыш по сравнению с максимально возможным значением.

3. Критерии выбора оптимальных стратегий игроков

3.1. Критерии оптимальности стратегий для первого этапа игры

Оптимальную стратегию администратора безопасности, который осмотритель-но играет и стремится минимизировать свой возможный максимальный ущерб, в простейшем случае можно найти из условия **минимакса** [1]. В таком случае критерий выбора оптимальной стратегии x_{i0} для администратора состоит в том, чтобы из его платёжной матрицы A взять следующее число:

$$W_{i0}(A) = \min_i \max_j a_{ij}. \quad (1)$$

Если атакующий злоумышленник также ориентируется на самые неблагоприятные условия, то он стремится максимизировать свой возможный минимальный выигрыш. Поэтому его оптимальную стратегию в простейшем случае можно найти из условия **максимина** [1]. Тогда критерий выбора оптимальной стратегии y_{j0} для злоумышленника состоит в том, чтобы из его платёжной матрицы B взять следующее число:

$$W_{j0}(B) = \max_j \min_i b_{ij}. \quad (2)$$

Минимаксная (1) и максиминная (2) стратегии игроков уместны в тех случаях, когда они не столько хотят выиграть, сколько не хотят проиграть. Хотя администратор безопасности и атакующий злоумышленник могут выбрать для себя и другие критерии для подбора наиболее оптимальных для них наборов программных средств, с помощью которых они смогут реализовать свои лучшие стратегии игры [3].

3.2. Критерии оптимальности стратегий для второго этапа игры

Необходимость в проведении второго этапа биматричной игры возникает, если на первом этапе найдено несколько наилучших стратегий игроков, имеющих одинаковую цену игры. В таком случае можно предположить, что на втором этапе игры второй игрок выбирает свою стратегию случайным образом, и воспользоваться теорией статистических решений [4]. Тогда среди найденных стратегий поиск наилучшей можно проводить, используя критерии оптимальности, основанные на оценке

степени удачности многократных применений стратегии, т. е. на оценке величины риска. Если оба игрока играют осторожно, то в простейшем случае для поиска их лучших стратегий наиболее подходит **критерий Сэвиджа (минимального максимального риска)** [4].

Согласно критерию Сэвиджа для администратора безопасности в каждой строке его матрицы рисков R определяется максимальный элемент, лучшей считается та стратегия, для которой этот результат наименьший. Таким образом, на втором этапе биматричной игры критерий выбора оптимальной стратегии x_{i_0} для администратора безопасности состоит в том, чтобы из его матрицы рисков R взять следующее число:

$$W_{i_0}(R) = \min_i \max_j r_{ij}. \quad (3)$$

Аналогично для злоумышленника в каждом столбце его матрицы рисков V определяется максимальный элемент и лучшей считается та стратегия, для которой этот результат наименьший. Критерий выбора оптимальной стратегии y_{j_0} для злоумышленника на втором этапе биматричной игры состоит в том, чтобы из его матрицы рисков V взять следующее число:

$$W_{j_0}(V) = \min_j \max_i v_{ij}. \quad (4)$$

Критерий минимаксного риска Сэвиджа предполагает, что оптимальной является та стратегия игрока, при которой величина риска для него в наихудшем случае минимальна.

4. Проведение второго этапа биматричной игры для уточнения лучших стратегий игроков

Рассмотрим на простом примере второй этап решения биматричной игры для случая, когда на первом этапе были найдены лучшие стратегии x_1, x_2, x_3 для администратора безопасности и y_1, y_2, y_3 – для злоумышленника. Составим из этих стратегий новые платежные матрицы A и B (см. табл. 3 и 4).

Таблица 3. Платёжная матрица A

	y_1	y_2	y_3
x_1	4	10	4
x_2	10	5	7
x_3	3	5	10

Таблица 4. Платёжная матрица B

	y_1	y_2	y_3
x_1	4	9	2
x_2	8	2	3
x_3	2	3	7

Будем считать, что администратор безопасности и атакующий злоумышленник крайне осторожны. Тогда их оптимальные стратегии можно найти из условий минимакса (1) и максимина (2) соответственно. При использовании этих критериев оцениваются значения ячеек платёжных матриц A и B , но, как видно, для рассматриваемой биматричной игры этого оказывается недостаточно для однозначного определения лучших стратегий игроков.

Для того чтобы составить матрицу рисков R администратора безопасности, нужно в каждом столбце его платёжной матрицы A найти наименьшее значение, по очереди вычесть его из всех значений в данном столбце и результат записать в те же позиции (см. табл. 5). Матрица рисков V злоумышленника строится по строкам его платёжной матрицы B . В каждой строке нужно найти наибольшее значение, из этого значения по очереди вычесть все значения в данной строке и результат записать в те же позиции (см. табл. 6).

Таблица 5. Матрица рисков R

	y_1	y_2	y_3
x_1	1	5	0
x_2	7	0	3
x_3	0	0	6

Таблица 6. Матрица рисков V

	y_1	y_2	y_3
x_1	5	0	7
x_2	0	6	5
x_3	5	4	0

После этого в каждой строке матрицы рисков R администратора безопасности и каждом столбце матрицы рисков V злоумышленника нужно определить максимальный элемент. Лучшей по критерию Сэвиджа считается та стратегия, для которой этот результат будет наименьшим. В нашем случае, согласно формулам (3) и (4), получается $W_{i_0=1}(R) = 5$ и $W_{j_0=1}(V) = 5$, поэтому лучшими стратегиями для администратора безопасности и атакующего злоумышленника будут x_1 и y_1 соответственно. Таким образом, если администратор безопасности и злоумышленник выберут стратегии x_1 и y_1 , то цена игры для каждого них будет по 4 у.е. (см. табл. 3 и 4). Если же выбрать менее удачные согласно критерию Сэвиджа стратегии, например x_2 и y_3 , то можно убедиться, что оба игрока сыграют хуже, так как цена игры для администратора безопасности составит 7 у.е. (большой ущерб), а для злоумышленника – 3 у.е. (меньшая прибыль).

5. Заключение

В статье продемонстрировано преимущество решения биматричной игры между администратором безопасности и атакующим злоумышленником в два этапа. Возможность проведения второго этапа игры в большинстве случаев позволяет выявить наиболее выигрышные стратегии игроков среди нескольких с одинаковой ценой игры, найденных на первом этапе. На основе предложенного подхода можно создать программное приложение, которое позволит администратору анализировать и оптимизировать подбор программных средств для построения системы безопасности компании.

Литература

1. Гуц А.К., Вахний Т.В. Теория игр и защита компьютерных систем: учебное пособие. Омск: Изд-во ОмГУ, 2013.

2. Вахний Т.В., Гуц А.К., Константинов В.В. Программное приложение для выбора оптимального набора средств защиты компьютерной информации на основе теории игр // Вестник Омского университета. 2013. № 4 (70). С. 201–206.
3. Вахний Т.В., Вахний С.В. Решение биматричной игры с применением различных критериев для выбора стратегий администратора безопасности и злоумышленника // Математические структуры и моделирование. 2023. № 3 (67). С. 111–120.
4. Шевченко Д.В. Методы принятия управленческих решений: задания и метод. указания для выполнения расчёт.-граф. работы. Казань: Познание, 2014.

OPTIMIZING THE CHOICE OF A STRATEGY TO PROTECT AGAINST MALICIOUS ATTACKS BASED ON A RISK-BASED BIMATRIC GAME SOLUTION

T.V. Vakhniy

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

S.V. Vakhniy

Student, e-mail: vakhniysv@mail.ru

Dostoevsky Omsk State University, Omsk, Russia

Abstract. The article describes the solution of a bimatrix game between a security administrator and an attacking attacker in two stages. At the first stage, it is proposed to apply optimality criteria that take into account the price of the game, and at the second stage, criteria that take into account the risks of the players. This approach helps to find the best strategy to protect against malicious attacks among strategies with the same price of the game.

Keywords: computer system, digitalization, cybersecurity, bimatrix game, optimal strategy.

Дата поступления в редакцию: 27.02.2024

КИБЕРАТАКИ: НЕКОТОРЫЕ ПОДХОДЫ К СИСТЕМНОМУ АНАЛИЗУ

А.И. Горев

к.ю.н., доцент, e-mail: gorevai@omsu.ru

Е.Г. Горева

к.ф.-м.н., e-mail: gorevaeg@omsu.ru

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

Аннотация. Отсутствие единого подхода к пониманию и реагированию на кибератаки и инциденты между производителями программного обеспечения в сфере защиты информации, пользователями и государственными структурами, курирующими сферу кибербезопасности, приводит к невозможности взаимодействия и совместного противодействия. Рассмотрены различные подходы к классификации кибератак. Сделана попытка систематизации и анализа существующих угроз в виде атак на сетевую информационную инфраструктуру.

Ключевые слова: компьютерная атака, информационная инфраструктура, компьютерный инцидент, структура, классификация.

Федеральный закон от 26.07.2017 № 187-ФЗ «О безопасности критической информационной инфраструктуры Российской Федерации» (далее – ФЗ о КИИ) определил начало нового этапа существования государства в эпоху информационного общества. До принятия закона задачи борьбы с нелегитимными действиями в сети возлагались на собственника информационных ресурсов, что ставило под сомнение успешность противодействия преступлениям в виртуальном пространстве. Это объяснялось многими факторами, в числе которых можно выделить отсутствие обмена опытом по противодействию преступному посягательству между собственниками информационных ресурсов, незаинтересованность собственников ресурсов в их защите от утечек информации и др. ФЗ о КИИ определил необходимость создания централизованного формата сбора, обработки и хранения информации об инцидентах информационной безопасности на объектах критической инфраструктуры.

Выделим два термина, определённых законодателем:

- компьютерная атака – целенаправленное воздействие программных и (или) программно-аппаратных средств на объекты критической информационной инфраструктуры (далее – КИИ), сети электросвязи, используемые для организации взаимодействия таких объектов, в целях нарушения и (или) прекращения их функционирования и (или) создания угрозы безопасности обрабатываемой такими объектами информации;
- компьютерный инцидент – факт нарушения и (или) прекращения функционирования объекта КИИ, сети электросвязи, используемой для организации

взаимодействия таких объектов, и (или) нарушения безопасности обрабатываемой таким объектом информации, в том числе произошедший в результате компьютерной атаки.

Интерес к терминологии объясним необходимостью однозначного понимания применяемых при взаимодействии аббревиатур, выражений, терминов. Формализация запросов и сообщений, регламентированная Национальным координационным центром по компьютерным инцидентам (далее – НКЦКИ), не может быть реализована без предварительного терминологического согласования. В настоящее время в сфере кибербезопасности нет единого подхода в данном вопросе, что обязательно приведёт к непониманию и коллизиям в правоприменительной практике взаимодействия с НКЦКИ. В формализованном сообщении о компьютерном инциденте субъект КИИ обязан сообщить о случившемся. Но если это сообщение будет формализовано только формально, т. е. при наличии структуры сообщения с выделенными полями об объекте КИИ, дате и времени инцидента и пр., поле с описанием зафиксированного события будет заполняться пользователем произвольно, то задачи систематизации, возложенные на НКЦКИ, не смогут быть реализованы в приемлемое время. Исходя из обнародованной статистики только зафиксированных кибератак, можно утверждать, что ежедневное количество сообщений о компьютерных инцидентах в НКЦКИ будет составлять десятки тысяч, что делает задачу обобщения практики и выдачу рекомендаций невыполнимой без предварительной формализации представления информации.

Отметим, что в сфере кибербезопасности нет единого сложившегося подхода к оценке инцидентов, что объясняется естественностью процесса развития любой сферы человеческой деятельности: появление новых технологий даёт развитие новым общественным отношениям, что, в свою очередь, требует своего урегулирования только после практических наработок и накопленного опыта.

В настоящее время можно отметить существование различных подходов к классификациям кибератак даже среди крупных производителей в сфере программного обеспечения (далее – ПО) защиты информации. Так, «Лаборатория Касперского» выделяет следующие виды: вредоносное ПО (далее – ВПО), распределённые атаки типа «отказ в обслуживании» (DdoS), фишинг, атаки с использованием SQL-инъекций, межсайтовый скриптинг (XSS) и ботнеты [1].

Positive Technologies предлагает следующую классификацию: фишинг, брутфорс-атака, ВПО, атака Drive-By Download, SQL-инъекции, атака «человек посередине» (Man-in-the-Middle, MITM), атаки типа «отказ в обслуживании» (Denial of Service, DoS) [2].

PVS-Studio выделяет среди основных видов кибератак: применение вредоносных программ; DdoS-атаки; фишинг; организация ботнетов; SQL-инъекция (внедрение кода SQL); XSS (межсайтовый скриптинг или применение межсайтовых сценариев); использование программы-вымогателя или шантажиста [3].

Подобная ситуация и в иностранных источниках. Так, М. Шиванандхан выделяет следующие виды атак: атаки типа «человек посередине» (MITM), фишинг и ВПО, скрытые загрузки во время перехода по гиперссылкам, ботнет-атаки, атаки социальной инженерии, атаки с использованием SQL-инъекции, вредоносные атаки, атаки с использованием межсайтовых скриптов (XSS), атаки с использованием паролей,

атаки типа «отказ в обслуживании» (DoS), распределённые атаки типа «отказ в обслуживании» (DdoS), внутренние атаки и утечка данных, атаки с криптоджекингом: деньги для вредоносных майнеров [4].

Сложность обобщения состоит в том, что благодаря современному состоянию информационного пространства право голоса имеют не только традиционные СМИ и компании, занимающиеся профессиональной деятельностью в данной сфере, но и любой блогер, желающий высказаться по вопросу. При этом уровень профессиональных компетенций никем не определён и не оценивается, благодаря чему в инфосфере создаётся высокий уровень «информационного шума». Как один из многих негативных примеров можно привести блог М. Кульгина, автор которого предлагает выделить 54 (!) вида атак, смешивая и многократно повторяя схожие составы [5].

Уведомления о компьютерной атаке, которые можно экспортировать в НКЦКИ, включают следующие виды: DdoS-атака, неудачные попытки авторизации, попытки внедрения ВПО, попытки эксплуатации уязвимости, публикация мошеннической информации, сетевое сканирование, социальная инженерия. При этом допустимые категории и типы инцидентов НКЦКИ, которые можно экспортировать в НКЦКИ, включают: вовлечение контролируемого ресурса в инфраструктуру ВПО, замедление работы ресурса в результате DdoS-атаки, заражение ВПО, захват сетевого трафика, использование контролируемого ресурса для фишинга, компрометация учётной записи, несанкционированное изменение информации, несанкционированное разглашение информации, публикация на ресурсе запрещённой законодательством РФ информации, рассылка спам-сообщений с контролируемого ресурса, успешная эксплуатация уязвимости [6].

В табл. 1 сведены сравнительные классификации атак производителей ПО защиты информации и формализованного запроса НКЦКИ.

Приведённый анализ показывает различие в подходах классификации у производителей ПО защиты информации и регламентирующим данный процесс государственным органом в лице НКЦКИ. Но менее объяснимым является то, что при определении перечня компьютерных атак отечественные производители ПО защиты информации не выделяют наиболее опасные виды, такие как АРТ- и LoTL-атаки. При этом описание данных видов атак присутствует в блогах указанных фирм [7, 8].

С другой стороны, отдельным видом кибератак выделены ботнеты, хотя ботнет определяется как единая сеть компьютеров, которыми можно управлять удалённо и контроль над которыми получен при использовании специальных троянских программ [9]. Таким образом, ботнет является инструментом для осуществления различных действий, в том числе атак. Вызывают нарекания и некоторые другие выделенные категории. Но особое внимание следует обратить на то, что оценки кибератак со стороны производителей ПО защиты информации сильно отличаются от классификации, предлагаемой НКЦКИ. В некоторых случаях однотипные атаки имеют различные наименования, что также недопустимо. Данное несоответствие существенным образом негативно скажется на практике регистрации кибератак субъектами КИИ, передаче данных об атаках в НКЦКИ, последующем обобщении зафиксированной информации и выдачу рекомендаций конечным пользователям.

Исходя из этого, считаем необходимым обратить особое внимание на классификацию кибератак. Основываясь на анализе источников, считаем необходимым

Таблица 1. Сравнительные классификации атак

Тип атаки	Лаборатория Касперского	Positive Technologies	PVS-Studio	НКЦКИ
Вредоносное ПО	+	+	+	+
Распределенные атаки				
«отказ в обслуживании» DdoS	+	–	+	+
«Отказ в обслуживании» DoS	–	+	–	–
Фишинг (социальная инженерия)	+	+	+	+
Использование SQL-инъекций	+	+	+	–
Межсайтовый скриптинг (XSS)	+	+	–	–
Ботнеты	+	–	+	–
Брутфорс-атака	–	+	–	–
Drive-By Download	–	+	–	–
«Человек посередине», MITM	–	+	–	–
Программа-вымогатель или шантажист	–	–	+	–
Неудачные попытки авторизации	–	–	–	+
Попытки эксплуатации уязвимости	–	–	–	+
Публикация мошеннической информации	–	–	–	+
Сетевое сканирование	–	–	–	+

выделить следующие категории:

- Вредоносное ПО – представляет нецеленаправленные атаки на широкую аудиторию, например, пользователей сети «Интернет». «Программа-вымогатель или шантажист», выделенные в таблице, представляют частный случай данной категории.
- Распределённые атаки «отказ в обслуживании» DdoS – целевые атаки, направленные на блокирование работы конкретного сетевого ресурса. Атака может быть реализована следующими тремя механизмами: переполнение канала связи, т. е. различные типы флуда, использование уязвимости стека сетевых протоколов и атаки на прикладной уровень [10]. DoS является упрощённой разновидностью атаки DdoS. Отличительной особенностью является явное проявление момента атаки.
- АРТ-атаки – advanced persistent threat (расширенные, таргетированные постоянные угрозы) – хорошо организованная, тщательно спланированная кибер-

атака, направленная на конкретную компанию или целую отрасль. Отличительной особенностью является использование уязвимостей нулевого дня. Продолжительность атаки может составлять несколько месяцев. Атака «попытка эксплуатации уязвимости», выделенная НКЦКИ, является разновидностью АРТ-атаки на неисправленную администратором системы уязвимость.

- LoTL-атаки (Living off the Land – атаки «подножным кормом») – метод использования легитимного ПО для осуществления деструктивных действий, таких как: получение неправомерного доступа, повышение привилегий, отправка и получение файлов, изменение системных настроек и т. д. Популярными инструментами являются PowerShell, WMI, CMD, CLI, BASH и др. Отличительной особенностью является отсутствие «цифрового следа», поскольку используются инструменты операционной системы (далее – ОС).
- Атаки «человек посередине», MITM – тип кибер-атаки, при котором злоумышленник перехватывает передачу данных путём прослушивания или притворяясь легальным участником [11]. Подготовка к атаке проводится с помощью сниффера пакетов – прикладной программы, которая использует сетевую карту в режиме promiscuous mode (в этом режиме все пакеты, полученные по физическим каналам, сетевой адаптер отправляет приложению для обработки). Сниффер перехватывает все сетевые пакеты, которые передаются через определённый домен. Некоторые исследователи выделяют сниффер в самостоятельный вид атаки, однако снифферы работают в сетях на законном основании для диагностики неисправностей и анализа трафика.
- Атака «ускоренная передача данных» – захват канала передачи данных, при котором работа конкурирующих TCP-соединений в том же коммуникационном канале блокируется, поскольку из-за резко возросшей интенсивности трафика другие соединения диагностируют состояние затора и принимают соответствующие меры по уменьшению скорости передачи данных, фактически освобождая канал для злоумышленника. Отличительной особенностью является легитимность действий злоумышленника, который используя уязвимости протокола обмена может повысить скорость своего обмена с сервером за счёт других пользователей. Отличительной особенностью является отсутствие необходимости технических знаний. Данный вид атаки можно рассматривать как легитимную DoS-атаку.
- Фишинг (социальная инженерия) – совокупность методов, позволяющих обмануть пользователя и заставить его раскрыть свой пароль, номер кредитной карты и другую конфиденциальную информацию [12]. Атака «Drive-By Download» часто является следствием фишинга, поскольку происходит при обращении пользователя к «заражённому» сайту. Публикация мошеннической информации для обмана пользователя является разновидностью фишинга. Отличительной особенностью является отсутствие необходимости технических знаний. Фишинговые мошенники не пытаются воспользоваться уязвимостями в ОС устройства, они прибегают к методам социальной инженерии.
- Неудачные попытки авторизации (брутфорс-атака, Brute-force) – попытка подобрать пароль или ключ шифрования. Метод заключается в последовательном переборе разных комбинаций символов до тех пор, пока одна из них не

подойдёт. Поиск пароля ведётся с использованием специальных программ и сервисов [13].

- Использование SQL-инъекций – уязвимость, которая позволяет атакующему использовать фрагмент вредоносного кода на языке структурированных запросов (SQL) для манипулирования базой данных и получения доступа к потенциально ценной информации. Атаки на основе таких уязвимостей – одни из самых распространённых и опасных: они могут быть нацелены на любое веб-приложение или веб-сайт, которые взаимодействуют с базой данных SQL [14].
- Межсайтовый скриптинг (XSS) – атаки с внедрением вредоносного кода на доверенные веб-сайты, в процессе которой происходит внедрение вредоносных скриптов в контент веб-сайта. Эти скрипты включаются в динамический контент, отображаемый в браузере жертвы, и браузер выполняет их. В результате вредоносные скрипты могут получить доступ к конфиденциальной информации, сохранённой браузером и используемой на этом сайте. Отличительной особенностью является то, что атака нацелена не на приложение – риску подвергаются пользователи веб-приложения [15].
- Сетевое сканирование – является возможным предвестником атаки. Выделяются два вида. Ping-сканирование – обнаружение «живых» узлов в сети с помощью широковещательной рассылки пакетов ICMP. Сканирование портов – пробные попытки подключения к внешним узлам, определяет доступные порты на узлах, на основе чего делается предположение о типе используемой ОС или конкретного приложения, запущенного на конечном узле [16].

Подводя итоги, можно отметить, что современное состояние безопасности информационной инфраструктуры существенным образом зависит от понимания угроз в сетевом пространстве. Многообразие компьютерных атак делает необходимым проведение системного анализа и исследования уязвимостей, на которые нацелены эти угрозы. Дальнейшие разработки систем безопасности без проведения системного анализа кибератак, принятия единой классификации и единого подхода к анализу инцидентов на базе НКЦКИ делает эту работу неэффективной.

Литература

1. Предотвращение кибератак. URL: <https://www.kaspersky.ru/resource-center/preemptive-safety/how-to-prevent-cyberattacks> (дата обращения: 10.01.2024).
2. Основные типы кибератак и способы борьбы с ними. URL: <https://www.securitylab.ru/analytics/535598.php> (дата обращения: 23.01.2023).
3. Кибератака. URL: <https://pvs-studio.ru/ru/blog/terms/6651/> (дата обращения: 11.03.2023).
4. 13 типов кибератак, о которых вам следует знать в 2023 году. URL: <https://www.freecodecamp.org/news/types-of-cyber-attacks-to-know/> (дата обращения: 12.09.2023).
5. 54 вида кибератак, о которых следует знать в 2024 году. URL: <https://zen.ru/a/ZUYNJUusBWYVvzz> (дата обращения: 04.11.2023).
6. Допустимые категории и типы инцидентов НКЦКИ. URL: <https://support.kaspersky.ru/help/KUMA/2.1/ru-RU/220462.htm> (дата обращения: 29.11.2023).

7. 5 признаков АРТ-атаки и советы по ее предотвращению. URL: <https://www.kaspersky.ru/resource-center/threats/advanced-persistent-threat> (дата обращения: 23.01.2024).
8. Living-off-the-land (LotL): скрытые атаки, уничтожающие целые организации. URL: <https://www.securitylab.ru/analytics/546134.php> (дата обращения: 23.01.2024).
9. Что такое ботнет? URL: <https://www.kaspersky.ru/resource-center/threats/botnet-attacks> (дата обращения: 23.01.2024).
10. Классификация DDoS: полное руководство по типам атак. URL: <https://ddos-guard.net/ru/info/blog-detail/classification-of-ddos-attacks-a-short-overview-of-modern-approaches> (дата обращения: 26.01.2023).
11. Что такое атака Man-in-the-Middle (MITM)? Определение и предотвращение. URL: <https://www.securitylab.ru/blog/company/PandaSecurityRus/351898.php> (дата обращения: 24.02.2022).
12. Как работает современный веб-фишинг. URL: <https://www.anti-malware.ru/How-Modern-Web-Phishing-Works> (дата обращения: 08.09.2020).
13. Что такое брутфорс. URL: <https://developers.sber.ru/help/business-development/what-is-a-brute-force> (дата обращения: 23.06.2023).
14. Что такое SQL-инъекция? Определение и описание. URL: <https://www.kaspersky.ru/resource-center/definitions/sql-injection> (дата обращения: 06.03.2022).
15. Методы предотвращения межсайтовых скриптингов (XSS). URL: <https://www.geeksforgeeks.org/cross-site-scripting-xss-prevention-techniques/> (дата обращения: 02.10.2021).
16. Проводим сканирование сети самостоятельно. URL: <https://www.securitylab.ru/blog/company/pt/349200.php> (дата обращения: 04.08.2021).

RULES FOR JOURNAL ARTICLES “MATHEMATICAL STRUCTURES AND MODELING”

A.I. Gorev

Ph.D. (Law), Associate Professor, e-mail: gorevai@omsu.ru

E.G. Goreva

Ph.D. (Phys.-Math.), e-mail: gorevaeg@omsu.ru

Dostoevsky Omsk State University, Omsk, Russia

Abstract. The lack of a unified approach to understanding and responding to cyber attacks and incidents between information security software manufacturers, users and government agencies in charge of cybersecurity leads to the impossibility of interaction and joint counteraction. Various approaches to the classification of cyber attacks are considered. An attempt is made to systematize and analyze existing threats in the form of attacks on the network information infrastructure.

Keywords: computer attack, information infrastructure, computer incident, structure, classification.

Дата поступления в редакцию: 01.03.2024

РАЗЛИЧНЫЕ СХЕМЫ И ПОДХОДЫ К ПРОЦЕССАМ ЖУРНАЛИРОВАНИЯ

В.Н. Семенихин

аспирант, e-mail: svladimir-99@mail.ru

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

Аннотация. Вычислительная среда требует надёжного и комплексного процесса для отслеживания и документирования действий пользователей для поддержания доверия к системе. Для этой цели используется журналирование. Однако журналы, которые создаются в процессе жизненного цикла информационной системы, уязвимы для множества атак, в том числе модификации журналов, возможности стирания журналов и раскрытия конфиденциальности пользователя. Для решения этих проблем создано множество схем. Каждая из них имеет свои особенности и свои взгляды на процессы журналирования. Дается обзор различных схем для защиты журналов, подходов к процессу журналирования.

Ключевые слова: схемы журналирования, схемы логирования, логирование.

Введение

Вычислительная среда требует надёжного и комплексного процесса для отслеживания и документирования действий пользователей, внутренних и сторонних сервисов для поддержания доверия к системе. Для этой цели используется журналирование. Требуется уточнить, что логирование и журналирование очень близкие понятия, которые описывают один и тот же процесс. Однако главное отличие этих процессов в их документированности. В случае журналирования все процессы и алгоритмы описаны в руководстве пользователя, тогда как в случае процесса логирования некоторый функционал может оставаться сокрытым. Далее будем говорить про подходы, которые имеют чётко описанную структуру, соответствующую понятию «журналирование». Благодаря данному процессу мы можем установить наблюдение за системой, действиями администраторов и пользователей. Однако в современном мире журналы уязвимы для множества атак, таких как подделка, модификация, возможность стирания журналов или раскрытие конфиденциальности пользователя. К этому следует добавить и действия администраторов, поскольку они имеют неограниченный доступ к журналам, могут создавать, модифицировать и даже удалять их. Защита журналов от вредоносных действий является крайне важной задачей их ведения. Существующие схемы имеют ряд ограничений, включая изменчивость, трудоёмкость вычислений, отсутствие семантики, также остро стоит вопрос верифицируемости. Целью этой статьи является знакомство с различными схемами для защиты журналов, с подходами к процессу журналирования.

1. Связанные проблемы и критерии оценки

В наши дни информационные технологии не перестают развиваться и совершенствоваться. Системы становятся всё более комплексными, усложняя как свою структуру, так и функционал. Данные, циркулирующие в таких системах, тоже выходят на новый уровень. Их объём стремительно растёт с развитием информационных систем, появляется неоднородность из-за различий между платформами, появляется больше работы, связанной с их безопасностью. Подобное целостное видение вызывает некоторые опасения, например какой уровень безопасности эти системы могут обеспечить? И как они обеспечивают и защищают конфиденциальность своих пользователей? Могут ли они в достаточной степени обеспечить целостность данных пользователей? Для удовлетворения нужд клиентов и заинтересованных сторон в информационных системах необходимы комплексные механизмы безопасности. Функциональность и производительность информационных систем, включая их устройства, можно отслеживать и исследовать путём анализа журналов, генерируемых взаимосвязанными устройствами, и обеспечивать централизованный мониторинг в информационных системах. Подотчётные журналы аудита могут обеспечить более высокий уровень доверия для поставщиков услуг и арендаторов. Доверие и управление доверием очень важно для сервиса, инфраструктуры, надёжности предоставляемых услуг и всей информационной системы в целом.

Также, говоря о журналировании, нельзя забывать о компьютерной криминалистике. Для успешного расследования инцидентов требуется защита журналов от несанкционированного доступа. В критериях оценки доверенных вычислительных систем [1] требования безопасности журналов аудита описываются следующим образом: «информация аудита должна выборочно храниться и защищаться, чтобы действия, влияющие на безопасность, можно было отследить до ответственной за них стороны» и «данные аудита должны быть защищены от изменения и несанкционированного уничтожения, а также позволять обнаруживать и расследовать нарушения безопасности». Здесь важно отметить, что результаты экспертизы любых сервисов зависят от состояния исходных журналов [2].

Администраторы в режиме непосредственного и удалённого доступа и поставщики услуг имеют полный доступ к соответствующим ресурсам и могут представлять серьёзную угрозу информационной безопасности. Защита ресурсов информационной системы является ключевым требованием для надёжной среды. Журналы аудита используются для мониторинга производительности, ресурсов, действий администраторов и пользователей. Журналирование помогает устранять проблемы в информационной системе. Эти журналы неоднородны в силу отсутствия требований к их реализации, поэтому часто структуры данных бывают непонятны вычислительным системам, поскольку им не хватает данных о семантике. Чтобы воспользоваться преимуществами журналирования и получить неизменяемую систему хранения с поддержкой гетерогенности, семантически обогащённую, и при этом иметь встроенную функцию обмена статистикой для заинтересованных сторон, необходимо решить следующие проблемы:

1. Обеспечение конфиденциальности, целостности, доступности.
2. Неоднородность журналов.

3. Проблема угроз злонамеренных пользователей.
4. Вопрос семантики.
5. Вопрос энергозависимости хранилищ журналов.

Для защиты целостности журналов уже проделана большая работа и до сих пор продолжают поиски наилучшего решения. В этой статье мы упомянем такие решения, как: BAF, BAFi и Fi-BAF, FssAgg, SecLaaS, SecLaaS-RW, BBox, SLOPPI, D-CAM, EmLog, Flogger, smartFIX, а также подходы на аппаратном уровне. Обеспечение безопасности (конфиденциальность, целостность, доступность), неизменяемости, семантики является критически важным вопросом схем журналирования.

2. Классические подходы и схемы

Для журналирования доступно множество различных решений. Решения для управления информацией о безопасности и событиях (SIEM) с открытым исходным кодом, такие как GFI Events Manager, Syslog-ng, Manage-Engine Log Storage and Analyser, LOGalyze, Splunk Enterprise и т. д., – вот несколько примеров таких решений. Стоит упомянуть различные DLP системы, которые тесно взаимодействуют с журналами и данными внутри информационной системы, а также помогают в расследовании инцидентов. Стоит также сказать про решения для анализа и поиска данных, такие как OpenSearch [3] или его отечественный аналог Arenaldata LogSearch [4]. Они представляют собой легкомасштабируемую систему для обеспечения быстрого доступа и реагирования на большие объёмы данных, независимо от их формата. Все эти решения способны хранить и анализировать журналы различных устройств. Системные администраторы имеют полный контроль над журналами в таких системах, и целостность хранилищ таких журналов объективно не может быть гарантирована. Обратим внимание на такой параметр, как неизменяемость данных. Для поддержания целостности журналов даже после компрометации системы исследователи предлагают разные решения. Рассмотрим некоторые из них.

2.1. Подходы на основе криптографии

В первом из примеров обеспечения безопасности данных используются асимметричные схемы шифрования. LogCrypt7 – асимметричная схема шифрования для обеспечения безопасности журналов, где журналы защищены шифрованием и доступны для публичной проверки. В этой схеме существуют дополнительные накладные расходы на шифрование, процессы дешифрования и сохранение ключей для этих типов шифрования.

Схемы на основе подписей с использованием шифрования, такие как BAF, BAFi, Fi-BAF и т. д., были предложены в [5–7]. Проблема данного решения в том, что журнал будет накапливать записи в течение длительного периода, а количество подписей, сохраняемых для проверки журнала, будет пропорционально расти. Это привело к новому решению на основе агрегирования. Схема FssAgg [8] была предложена для безопасной аутентификации на основе агрегирования этих подписей, чтобы защитить только ранее вошедшие в систему длинные сеансы. В схеме аутентификации FssAgg подписи с прямой защитой (MAC), сгенерированные одним и тем же под-

писывающим лицом, последовательно объединяются в одну совокупную подпись. Успешная проверка совокупной подписи эквивалентна проверке каждой подписи компонента. Неудачная проверка совокупности подписей подразумевает, что подпись хотя бы одной компоненты недействительна [9].

2.2. Подходы на основе облачных вычислений

В сценарии облачных вычислений для пользователей облака также предлагается ведение журнала как услуги. Эта схема также зависит от надёжности и добросовестности поставщиков облачных услуг, где арендаторы системы имеют возможность манипулировать этими журналами. Схема получения, хранения и анализа этих журналов через использование центрального сервера журналов SecLaaS предложена в [10] через использование облачных функций для хранения журналов. Но это дискуссионное решение, так как администраторы облака все ещё могут представлять угрозу. Эта система хранит данные виртуальных машин и предоставляет доступ криминалистам, обеспечивая конфиденциальность пользователей облака. Кроме того, SecLaaS защищает целостность журналов от нечестных следователей или поставщиков облачных услуг. К тому же реализация SecLaaS выполнена в OpenStack – популярной облачной платформе с открытым исходным кодом. Однако у этой схемы всё ещё остаётся проблема вседозволенности администраторов, что ставит целостность данных под сомнение.

В [11] авторы предложили SecLaaS-RW, используя «двустороннее нанесение водяных» знаков, где журналы хранятся в облаке в течение более длительного времени и аутентификация контента осуществляется при помощи нанесения обратимых водяных знаков. При таком подходе возможна только криминалистическая проверка подлинности контента, тогда как другие параметры безопасности отсутствуют.

Цифровой чёрный ящик (VBox) [12], обеспечивающий аутентичное архивирование в распределённых системах, создан для обеспечения подлинности и конфиденциальности журналов. Он основан на криптографии с открытым ключом. VBox использует стандартные примитивы для обеспечения подлинности записей во время передачи от устройств к источнику данных, а также во время хранения на источнике и извлечения аудиторами. В этом решении появляется дополнительная сущность, которая по сути становится единой точкой отказа всей системы.

Схема SLOPPI [13] сделана на основе шифрования для обеспечения целостности журналов и их соответствия заданной политике. В этом решении на журналирование обращают внимание с позиции администрирования и расследования инцидентов. Схема направлена на то, чтобы защитить журналы от злоумышленников, при этом оставив возможность администраторам выполнять свои базовые задачи. В этом решении обеспечение неизменности и семантики журналов отсутствует, нет защиты и от недобросовестности администраторов.

Анонимизация данных журнала путём переноса всей инфраструктуры в облачные хранилища, как предложено в [14], помогает обеспечить конфиденциальность данных, но другие требования безопасности данных при таком подходе отсутствуют.

Хенце с соавторами предложили структуру D-CAM [15] для безопасного журналирования в IoT. Предлагаемая структура позволяет осуществлять контроль и

управление из центра для защиты сети IoT от поставщика облачных услуг со злыми намерениями. Система определяет и регистрирует управляющие сообщения в резервных местах для проверки через различные шлюзы. Сообщения журнала проверки используются для выявления злонамеренного поведения, что помогает защитить облачный IoT от блокировки, вставки, удержания и изменения порядка сообщений.

2.3. Подходы на аппаратном уровне

Некоторые исследователи предложили аппаратные схемы безопасного ведения журнала. В [16] авторы представили хранение журналов в рамках концепции обеспечения целостности и неизменяемости с использованием сценариев однократной записи и многократного чтения (WORM). В качестве дополнения, блок управления чтением/записью отключает операции стирания в секторах, содержащих поля с уникальным форматированием, но разрешает такие операции в секторах, которые не содержат таких полей, чтобы разрешить тестирование калибровки и ведение списков дефектов носителей.

Защищённое от несанкционированного доступа хранение журналов с использованием Trusted Platform Module 2.0 (TPM) описано в [17]. Это обеспечивает решение практических проблем, в том числе обработку частого обновления журналов, экономию дискового пространства и эффективную проверку произвольного подмножества журналов. Для обеспечения гарантии защиты от несанкционированного доступа создан безопасный протокол регистрации.

Система EmLog [18] создана для защищённого хранения от несанкционированного доступа и защиты от модификаций. Данная схема разработана для ограниченных устройств с доверенной средой выполнения. Доверенная среда выполнения (TEE) – безопасная область главного процессора. Она помогает защитить код и данные, загруженные в него, с точки зрения конфиденциальности и целостности. Целостность данных предотвращает изменение данных неавторизованными объектами за пределами TEE, а целостность кода предотвращает замену или изменение кода в TEE неавторизованными объектами.

Все рассмотренные выше аппаратные схемы ограничены с точки зрения ресурсов и обработки. Их мощность, стоимость и доступность не подходят для более крупных сетей с множеством вычислительных блоков. Аналогично, другие неаппаратные схемы не могут в полной мере обеспечить неизменность, семантику и распределение статистики журналов.

2.4. Подходы на основе файловых систем

Flogger [19] передаёт информацию из пространств ядра как виртуальных машин (VM), так и физических машин (PM) в облаке, обеспечивая тем самым полную прозрачность всего ландшафта данных в облаке. С помощью Flogger над ним можно построить сервисы, чтобы предоставить облачным провайдерам, конечным пользователям и регулирующим органам соответствующую информацию о происхождении данных. Схема позволяет конечному пользователю отслеживать, был ли «тронут» его файл неавторизованным пользователем.

Защищённая пересылка журнала и подпись только для добавления (LogFAS) [20] достигает наиболее желательных свойств как симметричных схем, так и схем на основе PKC. LogFAS может создавать публично проверяемые подписи с прямой защитой и ограниченной возможностью добавления без необходимости какой-либо онлайн-поддержки доверенного сервера или фактора времени.

2.5. Другие подходы к журналированию

Также стоит отметить использование структур данных «только для добавления», представленных в схеме Balloons [21]. Данная схема хранения записей журнала заключается в подходе, когда изначально доверенный автор хранит данные на недоверенном сервере или хранилище, а пользователи могут запрашивать доступ к ней или различным её версиям. Описанная структура называется аутентифицируемой, в том смысле, что каждая операция, совершаемая с данными, будет зафиксирована (через механизм снимков) и доступна для публичной проверки. Данная схема предназначена для смягчения предположения о доверии для обеспечения прозрачности изменений, а также сохраняет конфиденциальность ведения журнала.

Интересный подход к журналированию на основе семантики smartFIX был предложен в ссылках [22, 23]. В этом решении документы классифицируются автоматически на основе методов свободной формы и форм-анализа. Соответствующие данные извлекаются с использованием разных методов для каждого типа документа, который проверяется и оценивается через базу данных сопоставления и другие сложные методы, основанные на предыдущих знаниях. Качество данных обеспечивается за счёт математических и логических проверок. Данные, которые точно распознаны, публикуются для прямого экспорта. Иные, неуверенно распознанные данные пересылаются для ручной проверки. По своей сути это решение, которое создано для того, чтобы получить данные независимо от источника и формата, в котором они представлены. Это решение полностью закрывает вопросы гетерогенности и семантики. К сожалению, безопасность, неизменяемость и другие базовые требования остаются вне поля этого решения.

Также следует упомянуть решения на основе блокчейна. В Scyt1 [24] используется технология, называемая неизменяемыми журналами, которая используется в решениях для электронного голосования. Это обеспечивает целостность, подлинность и неотказуемость созданных журналов, поэтому в случае каких-либо событий аудиторы могут использовать их для расследования инцидентов.

На основании этого исследования были выделены некоторые схемы, их сравнение представлено в табл. 1 и 2.

Заключение

Проблемы организации журналирования, связанные с обеспечением безопасности, целостности, конфиденциальности, гетерогенности, неизменяемости, существуют уже много лет. За это время проблемы были рассмотрены по-разному и с использованием разных технологий. Важно помнить, что угрозы по отношению к системе аудита никуда не исчезают, а компании всё больше оцифровывают свой

Таблица 1. Сравнение схем журналирования

Схема	Краткое описание и сильные стороны	Вопросы
LogCrypt	Публично верифицируемые журналы	Проблемы с производительностью
FssAgg	Последовательная агрегация подписей, основанная на длинных сессиях	Только для активных сессий
BAF	Система, основанная на подписях(MAC)	Проблемы с производительностью, проблемы с излишним потреблением памяти
Balloons	Структура данных для хранения на недоверенном хранилище	Вопросы в отношении потребления памяти
BBox	Решение в рамках распределённых вычислительных систем, сохраняется публичная верифицируемость	Проблемы с производительностью, единая точка отказа
LogFAS	Публичная верифицируемость	Проблемы с производительностью
smartFIX	Взгляд на процесс журналирования на основе семантики	Не уделено внимание остальным параметрам безопасности
BAF / FI-BAF	Агрегирование, уменьшение затрат на вычисление	Проблемы с производительностью в области криптографии, вопросы по распределению логов
SLOPPI	Централизованный центр журналирования	Отсутствует безопасность журналов
SecLaaS	Централизованный центр журналирования, дополнительный уровень защиты от злоумышленников, ориентирование на заданную политику	Вопросы целостности и неизменяемости остаются закрытыми не полностью, данные из журналов могут быть модифицированы или удалены администраторами
SecLaaS-RW	Добавление проверки подлинности, сильный механизм аутентификации	Не продумана работа в режиме реального времени и работа с хранилищами
D-CAM	Система основанная на технологии распределенного реестра (DLT)	Описана только конфигурация хранилища журналов

Таблица 2. Сравнение схем журналирования (продолжение). К – конфиденциальность, Ц – целостность, Д – доступность, Н – неизменяемость, Г – гетерогенность, С – семантика

Схема	К	Ц	Д	Н	Г	С
LogCrypt	+	+	-	+	-	-
FssAgg	-	+	-	+	-	-
BAF	-	+	-	+	-	-
BBox	+	+	-	+	-	-
LogFAS	-	+	-	+	-	-
smartFIX	-	-	-	-	+	+
BAF / FI-BAF	-	+	-	+	-	-
SLOPPI	-	+	-	+	-	-
SecLaaS	-	+	+	-	-	-
SecLaaS-RW	+	+	-	+	-	-
D-CAM	+	+	+	+	-	-

бизнес. Именно поэтому стоит обращать внимание на то, какие риски имеются у организации, какие методы и какие технологии стоит применять для того, чтобы избавиться от рисков или сделать их приемлемыми.

Литература

1. Qiu L., Zhang Y., Wang F., Kyung M., Mahajan H.R. Trusted Computer System Evaluation Criteria. National Computer Security Cente, 1985.
2. Ye F., Zheng Y., Fu X., Luo B., Du X., Guizani M. TamForen: a tamper-proof cloud forensic framework // Transactions on Emerging Telecommunications Technologies. 2020. Vol. 33, Iss. 4.
3. Opensearch. URL: <https://opensearch.org> (дата обращения: 05.01.2024).
4. Arenadata LogSearch. URL: <https://docs.arenadata.io/adls/index.html> (дата обращения: 05.01.2024).
5. Yavuz A.A., Ning P. Baf: an efficient publicly verifiable secure audit logging scheme for distributed systems // 2009 Annual Computer Security Applications Conference. 2009. P. 219–228.
6. Yavuz A.A., Ning P., Reiter M.K. BAF and FI-BAF: efficient and publicly verifiable cryptographic schemes for secure logging in resource-constrained systems // ACM Transactions on Information and System Security. Vol. 15, Iss. 2. Art. 9, P. 1–28.
7. Kampanakis P, Yavuz A.A. BAFi: a practical cryptographic secure audit logging scheme for digital forensics // Secure Community Network. 2015. Vol. 8, Iss. 17. P. 3180–3190.
8. Ma Di, Tsudik G. Forward-secure sequential aggregate authentication // IEEE. 2007. URL: <https://eprint.iacr.org/2007/052> (дата обращения: 02.01.2024).

9. Ma Di, Tsudik G. A New Approach to Secure Logging // Data and Applications Security XXII, 22nd Annual IFIP WG 11.3 Working Conference on Data and Applications Security, London, UK, July 13-16, 2008, Proceedings. Vol. 5094. P. 48–63.
10. Zawoad S, Dutta AK, Hasan R. SecLaaS: secure logging-as-a-service for cloud forensics // ASIA CCS '13. 2013. P. 219–230.
11. Khan A, Yaqoob A, Sarwar K, Tahir M, Ahmed M. Secure logging as a service using reversible watermarking // Procedia Computer Science. 2017. Vol. 110. P. 336–343.
12. Accorsi R. BBox: a distributed secure log architecture // EuroPKI. Public Key Infrastructures, Services and Applications. 2010. P. 109–124.
13. Von Eye F., Schmitz D., Hommel W. SLOPPI-A Framework for Secure Logging with Privacy Protection and Integrity // ICIMP. 2013.
14. Rajalakshmi J.R., Rathinraj M., Braveen M. Anonymizing log management process for secure logging in the cloud // International Conference on Circuits, Power and Computing Technologies. 2014. P. 1559–1564.
15. Henze M., Wolters B., Matzutt R., Zimmermann T., Wehrle K. Distributed configuration, authorization and management in the cloud-based internet of things // 2017 IEEE Trustcom/BigDataSE/ICCESS. 2017. P. 185–192.
16. Jaquette G.A., Jesionowski L.G., Kulakowski J.E., McDowell J.A. Low cost tamper-resistant method for write-once read many (WORM) storage. US Patent. 2001. US6272086B1.
17. Sinha A., Jia L., England P., Lorch J.R. Continuous tamper-proof logging using TPM 2.0 // Trust and Trustworthy Computing. 2014. Lecture Notes in Computer Science. Vol. 8564.
18. Shepherd C., Akram R.N., Markantonakis K. EmLog: tamper-resistant system logging for constrained devices with TEEs // Information Security Theory and Practice. 2017. Vol. 10741. P. 75–92.
19. Ko R.K., Jagadpramana P., Lee B.S. Flogger: a file-centric logger for monitoring file access and transfers within cloud computing environments // 2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications. 2011. P. 765–771.
20. Yavuz A.A., Ning P., Reiter M.K. Efficient, compromise resilient and append-only cryptographic schemes for secure audit logging // Financial Cryptography and Data Security. 2012. Vol. 7397. P. 148–163.
21. Pulls T., Peeters R. Balloon: a forward-secure append-only persistent authenticated data structure // Computer Security ESORICS. 2015. Vol. 9327. P. 622–641.
22. Forcher B., Agne S., Dengel A., Gillmann M., Roth-Berghofer T. Semantic logging: towards explanation-aware das // International Conference on Document Analysis and Recognition. 2011. P. 1140–1144.
23. Shafiq M.O. Semantically Formalized Logging and Advanced Analytics for Enhanced Monitoring and Management of Large-scale Applications: Doctoral thesis. Calgary, Canada: University of Calgary, 2015.
24. Cucurull J., Puiggali J. Distributed immutabilization of secure logs // Security and Trust Management. 2016. Vol. 9871. P. 122–137.

VARIOUS SCHEMES AND APPROACHES TO LOGGING PROCESSES**V.N. Semenikhin**

Ph.D. Student, e-mail: svladimir-99@mail.ru

Dostoevsky Omsk State University, Omsk, Russia

Abstract. The computing environment requires a reliable and comprehensive process for tracking and documenting user activity to maintain trust in the system. Logging is used for this purpose. However, logs that are created during the life cycle of an information system are vulnerable to a variety of attacks, including modification of logs, the possibility of erasing logs, and exposing user privacy. Many schemes have been created to solve these problems. Each of them has its own characteristics and its own views on logging processes. An overview of various schemes for protecting logs and approaches to the logging process is given.

Keywords: auditing, logging schemes, logging.

Дата поступления в редакцию: 14.05.2023

Романьков Виталий Анатольевич

02.02.1948 – 13.12.2023



13 декабря 2023 года скоропостижно, в поезде, на пути в Омск, умер замечательный омский математик, доктор физико-математических наук (1992), профессор (1995).

В.А. Романьков родился в Казахстане, в городе Усть-Каменогорске. Окончил механико-математический факультет Новосибирского государственного университета (1971) по специальности «Математика» и аспирантуру (1974), защитив кандидатскую диссертацию по алгебре и логике.

В 1974 году приехал на работу в Омский государственный университет в должности старшего преподавателя математического факультета.

С 1978 года старший, затем главный научный сотрудник лаборатории комбинаторных и вычислительных методов алгебры и логики Омского филиала Института математики им. С.Л. Соболева СО АН СССР. Продолжал работу по совместительству в ОмГУ доцентом, а с 1992 года профессором.

Заведующий кафедрой информационных систем. Читал курсы по алгебре, теории игр, исследованию операций, теории вероятностей и криптографии.

Научные интересы: теория разрешимых групп; группы автоморфизмов для свободных нильпотентных и разрешимых групп; с 1990-х годов – криптография.

Руководил аспирантами. Под его руководством защищены 12 кандидатских диссертаций и 2 PhD диссертации.

Виталий Анатольевич обладал даром уважительно относиться к праву другого человека иметь свой отличный взгляд на ту или иную ситуацию и поступать отличным образом в соответствии со своими убеждениями.

В молодости он как-то сказал, отвечая студентам на вопрос «Кем бы в жизни Вы хотели стать?»:

– Я хотел бы стать хорошим математиком.

Ему это удалось...

Авторам

Предоставляемые данные и документы

Автор предоставляет в редакцию:

- рукопись статьи в формате \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 \LaTeX и стиля msmb.cls, предоставляемого редакцией <http://msm.univer.omsk.su/files/msmb.zip>. Рекомендуется установить компилятор \MiKTeX , 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.

Запрещается использовать в формулах буквы русского алфавита. Если без них никак не обойтись, то следует использовать команду $\text{\mbox{...}}$.

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.

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 ("abcdefg" или "abcdefg").

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 редколлекцией 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. (Techn.)

к.т.н. = Ph.D. (Techn.)

Инженер-программист = Software Engineer

Старший/младший научный сотрудник = Senior/Junior Scientist Researcher

Электронная почта для отправки статей

dmtree1@ya.ru — зам. главного редактора (ответственный за выпуск) Д.Н. Лавров.

Научный журнал

Математические структуры И моделирование

№ 1 (69)

Главный редактор
А.К. Гуц

Зам. гл. ред., выпускающий редактор
Д.Н. Лавров

Зам. гл. ред., технический редактор
Н.Ф. Богаченко

Корректор:
Д.С. Нерозник

Перевод на английский язык в авторской редакции

Адрес научной редакции

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

Омский государственный университет

E-mail: aguts@mail.ru, dmtrl@ya.ru

Электронная версия журнала:

<http://msm.omsu.ru>

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



Подписано в печать 05.04.2024. Формат 60 × 84 1/8.
Усл. печ. л. 15,23. Тираж 25 экз. Заказ № 105.

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