



№1-2026

ISSN 1999-9429

ИЗВЕСТИЯ ЮФУ

ТЕХНИЧЕСКИЕ НАУКИ

- Алгоритмы обработки информации
- Анализ данных, моделирование и управление
- Электроника, нанотехнологии и приборостроение
- Машинное обучение и нейронные сети

ИЗВЕСТИЯ ЮФУ. ТЕХНИЧЕСКИЕ НАУКИ IZVESTIYA SFedU. ENGINEERING SCIENCES

Свидетельство о регистрации средства массовой информации

ПИ № ФС77-28889 от 12.07.2007

*Федеральная служба по надзору в сфере массовых коммуникаций, связи
и охраны культурного наследия*

Научно-технический и прикладной журнал

Издается с 1995 года, до середины 2007 года под названием «Известия ТРТУ»

Подписной индекс ПС704

№ 1 (249). 2026 г.

Журнал включен в «Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук».

Редакционный совет

Курейчик В.В. (гл. редактор); Кравченко Ю.А. (зам. гл. редактора); Бородинский И.М. (ученый секретарь); Абрамов С.М.; Агеев О.А.; Бабенко Л.К.; Боженюк А.В.; Борисов В.В.; Веселов Г.Е.; Гайдук А.Р.; Горбанёва О.И.; Еремеев А.П.; Зинченко Л.А.; Каляев И.А.; Касьянов А.О.; Коноплев Б.Г.; Коробейников А.Г.; Кравец А.Г.; Куповых Г.В.; Левин И.И.; Магид Е.А.; Массель Л.В.; Медведев М.Ю.; Мельник Э.В.; Никитов С.А.; Обуховец В.А.; Панич А.Е.; Петров В.В.; Пшихопов В.Х.; Редько В.Г.; Румянцев К.Е.; Сергеев Н.Е.; Середин Б.М.; Сидоркина И.Г.; Стемпковский А.Л.; Сухинов А.И.; Турулин И.И.; Тютиков В.В.; Угольницкий Г.А.; Целых А.Н.; Юханов Ю.В.

Учредитель Южный федеральный университет.

Издатель Южный федеральный университет.

Ответственные за выпуск: Клевцов С.И., Самойлов А.Н.

Технический редактор Ярошевич Н.В.

Оригинал-макет выполнен Ярошевич Н.В.

Дата выхода в свет 27.02. 2026 г. Формат 70×108 $\frac{1}{16}$. Бумага офсетная.

Офсетная печать. Усл. печ. л. – 25,9. Уч.-изд. л. – 19,7.

Заказ № 10349. Тираж 250 экз.

Адрес издателя: 344090, г. Ростов-на-Дону, пр. Стачки, 200/1, тел. 8(863)243-41-66.

Адрес типографии: Отпечатано в отделе полиграфической, корпоративной и сувенирной продукции Издательско-полиграфического комплекса КИБИ МЕДИА ЦЕНТРА ЮФУ. 344090, г. Ростов-на-Дону, пр. Стачки, 200/1, тел. 8(863)243-41-66.

Адрес редакции: 347922, г. Таганрог, ул. Чехова, 22, ЮФУ, тел. +7 (928) 909-57-82, e-mail: iborodyanskiy@sfedu.ru, <http://izv-tn.tti.sfedu.ru/>.

16+

Цена свободная

ISSN 1999-9429 (Print)

ISSN 2311-3103 (Online)

© Южный федеральный университет, 2026

СОДЕРЖАНИЕ

РАЗДЕЛ I. АЛГОРИТМЫ ОБРАБОТКИ ИНФОРМАЦИИ

В.И. Данильченко, В.В. Бова, М.М. Семенова, С.В. Игнатьева, М.Б. Шайлиев БИОИНСПИРИРОВАННЫЙ ПОДХОД ДЛЯ РЕШЕНИЯ ЗАДАЧИ ТРЕХМЕРНОЙ УПАКОВКИ.....	6
А.Р. Недвигин, Р.М. Синецкий АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ОЦЕНКИ ДОСТУПНОСТИ ТОВАРОВ В РАСПРЕДЕЛЕННЫХ СИСТЕМАХ ХРАНЕНИЯ НА ОСНОВЕ МЕТОДОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ	20
П.О. Никашина АЛГОРИТМ ОПРЕДЕЛЕНИЯ НЕЧЕТКОГО МНОЖЕСТВА СИЛЬНОЙ СВЯЗНОСТИ ПЕРИОДИЧЕСКОГО НЕЧЕТКОГО ГРАФА.....	33
М.В. Пикалов ВЛИЯНИЕ МЕТОДА ВЫБОРКИ НА ОЦЕНКУ ХАРАКТЕРИСТИК В АНАЛИЗЕ ЛАНДШАФТА ПОИСКА	42
М.О. Доброхвалов, А.Ю. Филатов, Е.А. Чегодаева НОВАЯ МЕТРИКА ВОСПРОИЗВОДИМОСТИ ДЛЯ СРАВНЕНИЯ КЛАССИФИКАТОРОВ ВРЕМЕННЫХ РЯДОВ.....	52

РАЗДЕЛ II. АНАЛИЗ ДАННЫХ, МОДЕЛИРОВАНИЕ И УПРАВЛЕНИЕ

К.Е. Румянцев, Д.А. Цыцорин МОДЕЛЬ ФОРМИРОВАНИЯ КЛЮЧЕВОЙ ПОСЛЕДОВАТЕЛЬНОСТИ В КОГЕРЕНТНОЙ ДВУХПОЛОСНОЙ ОПТИЧЕСКОЙ СИСТЕМЕ С МОДУЛЯЦИЕЙ ФАЗОМАНИПУЛИРОВАННЫМИ РАДИОСИГНАЛАМИ	65
А.Ю. Будко, Т.А. Гайда, З.А. Понимаш ОЦЕНКА ПРОСТРАНСТВЕННОГО ПОЛОЖЕНИЯ БОРТОВОЙ КАМЕРЫ ПУТЕМ СОПОСТАВЛЕНИЯ АЭРОФОТОСНИМКОВ И ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ.....	77
М.В. Грачев, Ю.Н. Паршин ПРОПУСКНАЯ СПОСОБНОСТЬ МОБИЛЬНОЙ МИМО СИСТЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ В НЕОДНОРОДНОЙ СРЕДЕ	86
Д.Ю. Зорькин, А.А. Богнюков, Т.Е. Кожанова РАЗРАБОТКА И ВНЕДРЕНИЕ КОРПОРАТИВНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ НА ПРОМЫШЛЕННОМ ПРЕДПРИЯТИИ ЗАВОДА «АвтоВАЗ».....	96
А.М. Пилипенко ПРИМЕНЕНИЕ ГИБРИДНЫХ МЕТОДОВ ЧИСЛЕННОГО РЕШЕНИЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ ДЛЯ АНАЛИЗА АВТОКОЛЕБАТЕЛЬНЫХ ЦЕПЕЙ С РАЗЛИЧНОЙ ДИНАМИКОЙ...	112
В.В. Гапочка, Е.Е. Полупанова ОБЗОР ТЕХНОЛОГИЙ СЕМАНТИЧЕСКОГО АНАЛИЗА И ИНТЕГРАЦИИ ГЕТЕРОГЕННЫХ ИНФОРМАЦИОННЫХ ПОТОКОВ В СИСТЕМАХ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ	122
Н.О. Лужевский, В.Ф. Лубенцов, Е.В. Лубенцова СИНТЕЗ РОБАСТНОЙ САУ С ДИНАМИЧЕСКИМ КОМПЕНСАТОРОМ И НАБЛЮДАТЕЛЕМ СОСТОЯНИЯ С КОРРЕКТИРУЮЩИМИ ОБРАТНЫМИ СВЯЗЯМИ НА ОСНОВЕ СИГМОИДНЫХ ФУНКЦИЙ	148
Д.Н. Богачева, О.В. Лукинова, А.А. Саломатин СЕМАНТИЧЕСКАЯ МОДЕЛЬ ВЛИЯНИЯ ФАКТОРОВ БЕЗОПАСНОСТИ ДЛЯ ОЦЕНКИ КИБЕРУСТОЙЧИВОСТИ ИНФОРМАЦИОННЫХ СИСТЕМ	164
Д.О. Ларин, Р.И. Захарченко, С.А. Диченко МАНДАТНАЯ АТРИБУТИВНО-РОЛЕВАЯ МОДЕЛЬ УПРАВЛЕНИЯ ДОСТУПОМ В КРУПНОМАСШТАБНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ	179

РАЗДЕЛ III. ЭЛЕКТРОНИКА, НАНОТЕХНОЛОГИИ И ПРИБОРОСТРОЕНИЕ

С.С. Быбин, Н.П. Дунаев, С.В. Кузьмин, А.Н. Морозов АНАЛИЗ ПРИЧИН ВОЗНИКНОВЕНИЯ ОШИБОК В АМПЛИТУДНО-ФАЗОВОМ РАСПРЕДЕЛЕНИИ ЛИНЕЙНЫХ ФАЗИРОВАННЫХ АНТЕННЫХ РЕШЕТОК И МЕТОДИКА ИХ УМЕНЬШЕНИЯ	192
А.И. Семенихин, Д.В. Семенихина, А.М. Зыкина МЕТАПОВЕРХНОСТИ С ОРБИТАЛЬНЫМ УГЛОВЫМ МОМЕНТОМ НА ОСНОВЕ ГЕКСАГОНАЛЬНЫХ ЯЧЕЕК ДЛЯ ШИРОКОПОЛОСНОГО СНИЖЕНИЯ РАССЕЯНИЯ	208
В.О. Игнатович, Н.Н. Кисель МОДЕЛИРОВАНИЕ И ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК МИКРОПОЛОСКОВОЙ МИМО-АНТЕННОЙ СИСТЕМЫ.....	218
А.В. Саенко, К.А. Козюменко, И.А. Шиховцов, Р.В. Томинов, В.А. Смирнов ИССЛЕДОВАНИЕ РЕЗИСТИВНОГО ПЕРЕКЛЮЧЕНИЯ ПРОЗРАЧНЫХ МЕМРИСТИВНЫХ СТРУКТУР НА ОСНОВЕ ОКСИДА ЦИНКА ДЛЯ МАШИННОГО ЗРЕНИЯ РОБОТОТЕХНИЧЕСКИХ СИСТЕМ.....	225

РАЗДЕЛ IV. МАШИННОЕ ОБУЧЕНИЕ И НЕЙРОННЫЕ СЕТИ

Б.В. Костров, С.И. Бабаев, А.И. Ефимов, В.Ю. Тарасова РАЗРАБОТКА ГИБРИДНОЙ АРХИТЕКТУРЫ СВЁРТОЧНОЙ НЕЙРОННОЙ СЕТИ С ИСПОЛЬЗОВАНИЕМ СПЕКТРАЛЬНЫХ ПРЕОБРАЗОВАНИЙ	235
В.Б. Савинов, Н.Н. Шушарина ОБЗОР МЕТОДОВ УЛУЧШЕНИЯ РАССУЖДЕНИЙ В БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЯХ	250
А.В. Бобряков, С.А. Прокопенко РАННЕЕ ВЫЯВЛЕНИЕ ПРОИЗВОДСТВЕННОГО БРАКА НА МЕЛКОСЕРИЙНЫХ ПРОИЗВОДСТВАХ С ИСПОЛЬЗОВАНИЕМ НЕЙРО-НЕЧЕТКИХ СИСТЕМ.....	270
В.Е. Бондарева, Т.С. Черноморова, А.В. Кривцун, А. Абдулкарём АЛГОРИТМ РАСПОЗНАВАНИЯ ЯЗЫКА ЖЕСТОВ С ИСПОЛЬЗОВАНИЕМ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ	284

CONTENT

SECTION I. INFORMATION PROCESSING ALGORITHMS

V.I. Danilchenko, V.V. Bova, M.M. Semenova, S.V. Ignateva, M.B. Shailiev	
A BIOINSPIRED APPROACH TO SOLVING THE PROBLEM OF 3D PACKAGING	6
A.R. Nedvigina, R.M. Sinetsky	
METHODOLOGICAL SUPPORT FOR ASSESSING THE AVAILABILITY OF GOODS IN DISTRIBUTED STORAGE BASED ON COMPUTER VISION METHODS	21
P.O. Nikashina	
ALGORITHM FOR DETERMINING A STRONGLY CONNECTED FUZZY SET OF A PERIODIC FUZZY GRAPH	34
M.V. Pikalov	
IMPACT OF SAMPLING TECHNIQUES ON EXPLORATORY LANDSCAPE ANALYSIS	42
M.O. Dobrokhvalov, A.Yu. Filatov, E.A. Chegodaeva	
A NEW REPRODUCIBILITY METRIC FOR COMPARING TIME SERIES CLASSIFIERS	53

SECTION II. DATA ANALYSIS, MODELING AND CONTROL

K.Y. Rumyantsev, D.A. Tsytosin	
MODEL OF KEY SEQUENCE GENERATION IN A COHERENT DUAL-BAND OPTICAL SYSTEM WITH PHASE-SHIFT KEYED RADIO SIGNAL MODULATION	65
A.Yu. Budko, T.A. Gayda, Z.A. Ponimash	
ESTIMATION OF THE SPATIAL POSITION OF AN ON-BOARD CAMERA BY COMPARING AERIAL IMAGES AND SATELLITE IMAGE DATA.....	78
M.V. Grachev, Yu.N. Parshin	
A CAPACITY OF MOBILE MIMO COMMUNICATION SYSTEM IN INHOMOGENEOUS MEDIUM	87
D.Yu. Zorkin, A.A. Bognyukov, T.E. Kozhanova	
DEVELOPMENT AND IMPLEMENTATION OF A CORPORATE INFORMATION SYSTEM AT THE AvtoVAZ INDUSTRIAL ENTERPRISE	97
A.M. Pilipenko	
APPLICATION OF HYBRID METHODS FOR NUMERICAL SOLVING OF ORDINARY DIFFERENTIAL EQUATIONS FOR ANALYSIS OF SELF-OSCILLATING CIRCUITS WITH VARIOUS DYNAMICS	113
V.V. Gapochka, E.E. Polupanova	
SEMANTIC ANALYSIS AND INTEGRATION OF HETEROGENEOUS INFORMATION STREAMS IN DECISION SUPPORT SYSTEMS: A TECHNOLOGY REVIEW	123
N.O. Luzhevsky, V.F. Lubentsov, E.V. Lubentsova	
SYNTHESIS OF A ROBUST ACS WITH A DYNAMIC COMPENSATOR AND A STATE OBSERVER WITH CORRECTIVE FEEDBACK BASED ON SIGMOID FUNCTIONS.....	148
D.N. Bogacheva, O.V. Lukinova, A.A. Salomatina	
SEMANTIC MODEL OF SECURITY FACTORS INFLUENCE FOR EVALUATING CYBER RESILIENCE OF INFORMATION SYSTEMS.....	165
D.O. Larin, R.I. Zaharchenko, S.A. Dichenko	
MANDATORY ROLE-CENTRIC ATTRIBUTE-BASED ACCESS CONTROL MODEL FOR LARGE-SCALE INFORMATION SYSTEMS.....	179

SECTION III. ELECTRONICS, NANOTECHNOLOGY AND INSTRUMENTATION

S.S. Bybin, N.P. Dunaev, S.V. Kuzmin, A.N. Morozov ANALYSIS OF THE CAUSES OF ERRORS IN THE AMPLITUDE-PHASE DISTRIBUTION OF LINEAR PHASED ANTENNA ARRAYS AND METHODS FOR THEIR REDUCTION	192
A.I. Semenikhin, D.V. Semenikhina, A.M. Zikina HEXAGONAL CELL-BASED ORBITAL ANGULAR MOMENTUM METASURFACES FOR BROADBAND SCATTERING REDUCTION	209
V.O. Ignatovich, N.N. Kisel MODELLING AND EXPERIMENTAL RESEARCH OF MICROSTRIP MIMO-ANTENNA SYSTEM'S CHARACTERISTICS	219
A.V. Saenko, K.A. Kozyumenko, I.A. Shikhovtsov, R.V. Tominov, V.A. Smirnov STUDY OF RESISTIVE SWITCHING OF TRANSPARENT ZINC OXIDE MEMRISTIVE STRUCTURES FOR MACHINE VISION OF ROBOTIC SYSTEMS	225

SECTION IV. MACHINE LEARNING AND NEURAL NETWORKS

B.V. Kostrov, S.I. Babaev, A.I. Efimov, V.Yu. Tarasova CONVOLUTIONAL NEURAL NETWORK HYBRID ARCHITECTURE DEVELOPMENT USING SPECTRAL TRANSFORMATIONS	235
V.B. Savinov, N.N. Shusharina A REVIEW OF METHODS FOR IMPROVING REASONING IN LARGE LANGUAGE MODELS	250
A.V. Bobryakov, S.A. Prokopenko EARLY DETECTION OF MANUFACTURING DEFECTS IN SMALL-SCALE PRODUCTION USING NEURO-FUZZY SYSTEMS	271
V.E. Bondareva, T.S. Chernomorova, A.V. Krivtsun, A. Abdulkarem ALGORITHM FOR SIGN LANGUAGE RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK	285

Раздел I. Алгоритмы обработки информации

УДК 004.89

DOI 10.18522/2311-3103-2026-1-6-20

В.И. Данильченко, В.В. Бова, М.М. Семенова, С.В. Игнатьева, М.Б. Шайлиев **БИОИНСПИРИРОВАННЫЙ ПОДХОД ДЛЯ РЕШЕНИЯ ЗАДАЧИ ТРЕХМЕРНОЙ** **УПАКОВКИ**

Рассматривается одна из важных комбинаторных задач оптимизации – задача трехмерной упаковки. Оптимизация трехмерной упаковки снижает затраты и повышает эффективность логистики, что делает ее актуальной для промышленности. В работе проанализированы классические подходы, такие как жадные алгоритмы и динамическое программирование, а также широко применяемые методы, включая эволюционные алгоритмы и локальный поиск. Анализ существующих методов, включая жадный поиск, динамическое программирование, эволюционные алгоритмы и локальный поиск, позволил выявить их ключевые характеристики и определить подходящие области применения. В контексте данного анализа представлен обзор ключевых методов, доминировавших в определенные исторические периоды. Анализ включает рассмотрение условий применения различных методов, их эффективности для определенных типов задач, а также их преимуществ и ограничений. Представлен многоуровневый алгоритм поиска, который объединяет преимущества традиционных и современных методов оптимизации. Многоуровневый алгоритм позволяет улучшить точность решения задачи упаковки за счет динамической настройки параметров. Разработан программный комплекс для решения задачи оптимизации трехмерной упаковки с использованием биоинспирированных алгоритмов. Проведен вычислительный эксперимент на тестовых примерах (бенчмарках). Качество упаковки, полученное, на основе разработанного комбинированного биоинспирированного алгоритма, в среднем на 7 % превосходит результаты упаковки, полученные с использованием известных алгоритмов, а время решения меньше от 7% до 25%, что говорит об эффективности предложенного подхода. Проведенные серии тестов и экспериментов позволили уточнить теоретические оценки временной сложности алгоритмов упаковки. В лучшем случае временная сложность алгоритмов $O(n^2)$, в худшем случае – $O(n^3)$.

Трехмерная упаковка; оптимизационные алгоритмы; эволюционные алгоритмы; многоуровневый алгоритм поиска; методы локального поиска.

V.I. Danilchenko, V.V. Bova, M.M. Semenova, S.V. Ignateva, M.B. Shailiev **A BIOINSPIRED APPROACH TO SOLVING THE PROBLEM OF 3D PACKAGING**

This article examines one of the most important combinatorial optimization problems – three-dimensional packaging. Optimizing three-dimensional packaging reduces costs and improves logistics efficiency, making it relevant for industry. This paper analyzes classical approaches such as greedy algorithms and dynamic programming, as well as widely used methods, including evolutionary algorithms and local search. An analysis of existing methods, including greedy search, dynamic programming, evolutionary algorithms, and local search, revealed their key characteristics and identified suitable areas of application. In the context of this analysis, an overview of the key methods that dominated during certain historical periods is presented. The analysis includes consideration of the application conditions of various methods, their effectiveness for specific types of problems, as well as their advantages and limitations. A multi-level search algorithm is presented that combines the advantages of traditional and modern optimization methods. This multi-level algorithm improves the accuracy of the packaging problem solution through dynamic parameter adjustment. A software package for solving the three-dimensional packaging optimization problem using bioinspired algorithms has been developed. A computational experiment was

conducted on test examples (benchmarks). The packing quality obtained using the developed combined bioinspired algorithm is, on average, 7% higher than the packing results obtained using known algorithms, while the solution time is 7% to 25% shorter, demonstrating the effectiveness of the proposed approach. A series of tests and experiments allowed us to refine theoretical estimates of the time complexity of packing algorithms. In the best case, the time complexity of the algorithms is $O(n^2)$, and in the worst case, $O(n^3)$.

Three-dimensional packing; optimization algorithms; evolutionary algorithms; multi-level search algorithm; local search methods.

Введение. Трехмерная упаковка минимизирует расходы и повышает эффективность использования производственных площадей. Однако их эффективность снижается с ростом объема данных и вычислительных требований.

Эволюционные алгоритмы предоставляют возможность нахождения глобально оптимальных решений, однако их реализация сопряжена с высокими вычислительными затратами, а также гибко адаптируются к различным задачам благодаря возможности настройки параметров и применения стохастических стратегий [1].

В данной статье представлен многоуровневый алгоритм поиска, который объединяет преимущества традиционных и современных методов оптимизации. Алгоритм позволяет улучшить точность решения задачи упаковки за счет динамической настройки параметров, что подтверждено результатами сравнительного анализа.

Актуальность проблемы трехмерной упаковки. Хотя задача оптимального размещения объектов в пространстве кажется простой, ее решение предполагает использование методов, таких как жадные алгоритмы, эволюционные стратегии и методы динамического программирования.

Основные аспекты проблемы:

- ♦ Экономический аспект. Оптимизация пространственного размещения объектов позволяет значительно сократить операционные расходы, включая затраты на логистику, хранение и производственные процессы. В производственных процессах рациональное распределение материалов минимизирует отходы и сокращает время сборки. Упаковка с высокой плотностью снижает количество транспортных единиц и ускоряет доставку.

- ♦ Технический прогресс. Современные технологии, особенно в микроэлектронике, предполагают интеграции большего числа компонентов на ограниченной площади, что делает стандарты трехмерной упаковки строгими.

- ♦ Экологический аспект. Оптимизация размещения объектов сокращает отходы упаковочных материалов и снижает негативное воздействие на окружающую среду.

- ♦ Вычислительная сложность. Задача трехмерной упаковки относится к NP-трудным задачам, что делает ее решение интеллектуально сложным и ресурсоемким. С увеличением числа элементов и их разнообразия возрастают вычислительные требования, стимулируя разработку новых методов.

Значимость трехмерной упаковки заключается в ее влиянии на ключевые показатели эффективности многих отраслей, включая производство, логистику и хранение [2]. В сфере микроэлектроники необходимость оптимального распределения компонентов возрастает с уменьшением размеров устройств.

Перспективными направлениями исследований являются разработка гибридных алгоритмов, которые объединяют методы локального поиска и эволюционные стратегии, а также применение искусственного интеллекта для автоматизации процесса оптимизации. Эти подходы обладают потенциалом для повышения точности и эффективности решений, а также расширения возможностей их применения в различных отраслях.

Постановка задачи. Основная проблема связана с необходимостью оптимизации процесса упаковки с учетом увеличения сложности и объема данных. На начальном этапе для тестирования алгоритма выбрана форма параллелепипеда, поскольку такой выбор минимизирует вычислительную сложность и стандартизирует условия экспериментов. В микроэлектронике параллелепипедная форма служит моделью для компонентов, таких как микросхемы и корпуса чипов, что способствует разработке алгоритмов их эффективного

размещения. Использование описанного подхода предоставляет возможность адаптации метода для работы с геометрически сложными объектами. Ключевые требования заключаются в разработке моделей и алгоритмов для оптимизации процесса упаковки, проведении оценки их эффективности и анализе возможностей их практического применения.

Исходными данными задачи упаковки являются область трехмерного пространства, список блоков и набор ограничений. Задача заключается в размещении блоков с учетом ограничений так, чтобы значение целевой функции стремилось к максимальному. Представленное количество блоков A , типов T , количеством N необходимо в соответствии с заданными ограничениями разместить в заданную трехмерную область, представленную в виде параллелепипеда M таким образом, чтобы значение целевой функции (ЦФ) стремилось к максимальному [3]. Результатом решения задачи трехмерной упаковки является план упаковки блоков, рассчитанное значение ЦФ, а также время, затраченное на решение задачи.

Критерий – это отношение суммарного объема упакованных блоков к объему трехмерной области упаковки.

Для записи параметров блока используется кортеж длиной 6 $\langle k_i, p_i, l_i, w_i, h_i, v_i \rangle$, где k_i – номер блока, p_i – ориентация (положение) блока в пространстве, l_i – длина блока, w_i – ширина блока, h_i – высота блока, v_i – объем блока [4].

Положение блока в трехмерном пространстве задается с помощью множества $S = \{s_i = \langle x_{1i}, y_{1i}, z_{1i} \rangle, \langle x_{2i}, y_{2i}, z_{2i} \rangle \mid i = 1, 2, \dots, n\}$, где $\langle x_{1i}, y_{1i}, z_{1i} \rangle$ – координаты угла, расположенного ближе всего к началу осей, а $\langle x_{2i}, y_{2i}, z_{2i} \rangle$ – координаты угла, расположенного дальше всего к началу осей [4].

Область трехмерного пространства представляется в следующем виде $M = \{L, W, H\}$, где L – длина области упаковки, W – ширина области упаковки, H – высота области упаковки. В дальнейшем данный подход подлежит расширению для применения к геометрическим формам повышенной сложности, где L – длина области упаковки, W – ширина области упаковки, H – высота области упаковки.

Для решения задачи трехмерной упаковки введем следующие ограничения: блоки размещаются исключительно внутри границ области трехмерной упаковки (формула 1) [4]:

$$\begin{cases} x_{1i} \geq 0; \\ y_{1i} \geq 0; \\ z_{1i} \geq 0; \\ x_{2i} \leq L_x; \\ y_{2i} \leq W_y; \\ z_{2i} \leq H_z. \end{cases} \quad (1)$$

2) общий объем V блоков не превышает объема V трехмерной области упаковки (формула 2):

$$\sum_{i=1}^n (l_i \cdot w_i \cdot h_i) \leq L_x \cdot W_y \cdot H_z. \quad (2)$$

3) Блоки не пересекаются друг с другом (формула 3):

$$\begin{aligned} & (x_{2i} \cdot x_{1j} \& y_{2i} \geq y_{1j} \& z_{2i} \geq z_{1j}) \vee (x_{2i} \cdot x_{1j} \& y_{2i} \geq y_{1j} \& z_{2i} \geq z_{1j}) \vee \\ & (x_{2i} \cdot x_{1j} \& y_{2i} \geq y_{1j} \& z_{2i} \geq z_{1j}) \vee (x_{2i} \cdot x_{1j} \& y_{2i} \geq y_{1j} \& z_{2i} \geq z_{1j}) \vee \\ & (x_{2i} \cdot x_{1j} \& y_{2i} \geq y_{1j} \& z_{2i} \geq z_{1j}) \vee (x_{2i} \cdot x_{1j} \& y_{2i} \geq y_{1j} \& z_{2i} \geq z_{1j}) \vee \\ & (x_{2i} \cdot x_{1j} \& y_{2i} \geq y_{1j} \& z_{2i} \geq z_{1j}) = 1 \forall i \leq n, j \leq (i \neq j). \end{aligned} \quad (3)$$

4) блоки должны опираться своим основанием на поверхность;

5) положение блока в пространстве допускается только перпендикулярным области упаковки.

Критерием задачи трехмерной упаковки является объем V , занимаемый упакованными блоками в трехмерном пространстве.

Для формулирования задачи и оценки ее результатов определим критерии и целевую функцию. Целевая функция представляется следующим образом (формула 5) [4]:

$$F = \frac{\sum V_{\text{бл}}}{V_{\text{кон}}} \rightarrow 1, \quad (5)$$

где $V_{\text{бл}}$ – объем блока, $V_{\text{кон}}$ – объем контейнера.

Целевая функция направлена на минимизацию пустот в упаковке, что способствует повышению плотности размещения и может снижать затраты на транспортировку и хранение. Максимальное значение функции, равное единице, соответствует полностью заполненному контейнеру [4]. Круг решения охватывает задачи, связанные с оптимизацией размещения объектов в контейнерах разного размера и формы, включая стандартные логистические контейнеры и специальные упаковочные модули. Предложенный метод может быть применим в логистике, производстве, микроэлектронике, а также в задачах проектирования, где требуется высокая плотность размещения. Предложенный подход может быть использован для решения задач различной сложности, от базовых случаев (параллелепипедные объекты) до высокотехнологичных сценариев.

Примером применения такой целевой функции является логистика, где минимизация пустот внутри транспортных контейнеров способствует сокращению количества рейсов, снижению затрат на топливо и повышению общей эффективности операций.

Данный подход предполагает минимизацию пустот в упаковочной области. Максимально эффективный результат решения задачи достигается, когда целевая функция (ЦФ) равна 1, что соответствует полному заполнению области упаковки.

Для решения задачи были введены следующие ограничения: блоки не выходят за пределы области упаковки, не пересекаются друг с другом, их объем не превышает объем области, а ориентация блоков остается перпендикулярной. В таблице представлены связи этих ограничений с ключевыми факторами

Таблица 1

Связи ограничений с ключевыми факторами

Ограничение	Связь с ключевыми факторами
Блоки размещаются исключительно внутри границ области трехмерной упаковки	Ограничение минимизирует вероятность ошибок при размещении, повышая точность логистических операций. Это связано с внедрением автоматизированных систем управления и использованием алгоритмов распознавания границ, что является примером технического прогресса.
Общий объем блоков не превышает объема упаковочной области	Поддерживает экологическую устойчивость за счет минимизации использования упаковочных материалов. Использование технологий перерасчета объема в реальном времени позволяет эффективно распределять ресурсы, что связано с экономическими и экологическими факторами
Блоки не пересекаются	Применение алгоритмов искусственного интеллекта, включая генетические и эвристические методы, исключает пересечения блоков, что способствует оптимизации размещения [5]. Эти технологии предотвращают ошибки и оптимизируют размещение блоков, что демонстрирует развитие автоматизации.
Блоки опираются на поверхность	Это ограничение увеличивает стабильность и качество упаковки, снижая риск повреждения продукции. Автоматическое распознавание опорных поверхностей улучшает логистические процессы и уменьшает количество брака, что положительно влияет на экономическую выгоду.
Положение блоков перпендикулярно	Упрощение расчетов и моделирования, достигаемое за счет перпендикулярного размещения блоков, снижает вычислительную сложность. Применение адаптивных алгоритмов обеспечивает эффективную интеграцию этого ограничения в автоматизированные системы, что способствует техническому прогрессу.

Методология исследования. Трехмерная упаковка является актуальной задачей для различных отраслей, учитывая ее сложность и важность.

В контексте данного анализа представлен обзор ключевых методов, доминировавших в определенные исторические периоды.

Анализ будет включать рассмотрение условий применения различных методов, их эффективности для определенных типов задач, а также их преимуществ и ограничений [6]. Такая аналитика позволяет получить представление о современном состоянии дисциплины в области трехмерной упаковки, а также выявить направления для последующих исследований:

а) Методы жадного поиска характеризуются высоким быстродействием, однако демонстрируют склонность к попаданию в локальные оптимумы.

б) Методы динамического программирования. Эти методы основаны на разбиении основной задачи на подзадачи. Для решения используются рекурсия и сохранение промежуточных результатов. Алгоритмы динамического программирования эффективны для определенных типов задач, но их сложность увеличивается пропорционально объему данных.

в) Эволюционные алгоритмы. Основаны на принципах естественной селекции и генетики. Через процессы мутации, кроссовера и селекции формируются новые поколения решений. Эволюционные методы позволяют искать глобальный оптимум, но требуют больше времени на вычисления.

г) Методы локального поиска, например 'имитация отжига', начинаются с произвольного решения и последовательно улучшают его за счет локальных модификаций. Алгоритмы локального поиска позволяют достичь баланса между вычислительной сложностью и качеством решения, однако их использование сопровождается риском нахождения локальных оптимумов.

В последние годы наблюдается тенденция к комбинированию различных методов с целью создания гибридных алгоритмов, которые пытаются объединить преимущества разных подходов и минимизировать их недостатки.

Для тестирования предложенных методов использовались параллелепипедные объекты с размерами, варьирующимися от 10x10x10 до 100x100x100 единиц. Выбранный диапазон размеров обеспечивает стандартизацию условий эксперимента.

Для симуляций использовались виртуальные параллелепипеды с плотностью и массой в диапазоне 0,5–2,5 кг. Эти значения были выбраны для отражения типичных характеристик грузов, используемых в промышленности и логистике, где диапазон массы и плотности объектов варьируется в зависимости от типа упаковки и условий хранения [7]. Такой подход способствует повышению реалистичности моделирования и проверке универсальности алгоритма. Предложенные алгоритмы тестировались в статических условиях, что позволило исключить воздействие температурных и внешних факторов, таких как перепады температуры, вибрации или воздействие влажности.

Разработан программный комплекс для решения задачи оптимизации трехмерной упаковки с использованием биоинспирированных алгоритмов. Основные функциональные возможности системы включают:

- ♦ Задание параметров задачи, таких как размеры и количество блоков, а также загрузка заранее подготовленных данных.
- ♦ Настройка параметров алгоритмов (рис. 1), включая:
 - параметры генетического алгоритма (размер популяции, число итераций, виды операторов и их вероятности);
 - параметры эволюционного алгоритма (оператор мутации, его вероятность, число итераций);
 - настройки сходимости и предельного времени выполнения

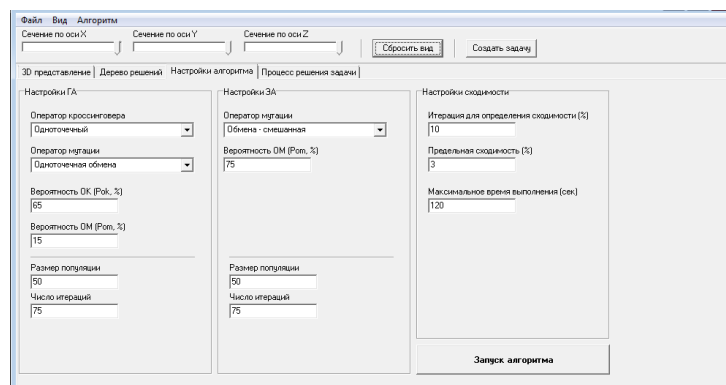


Рис. 1. Настройка параметров алгоритмов

- ◆ Визуализация процесса работы алгоритма в реальном времени с помощью:
 - графиков изменения целевой функции (рис. 2);
 - дерева решений, где отображаются промежуточные результаты (рис. 3).

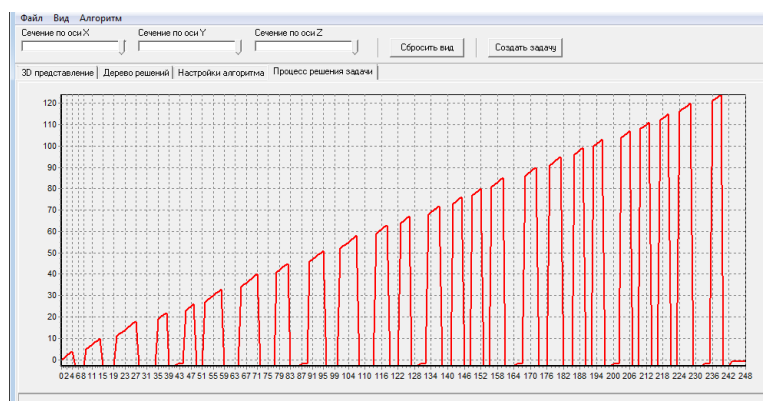


Рис. 2. График изменения целевой функции

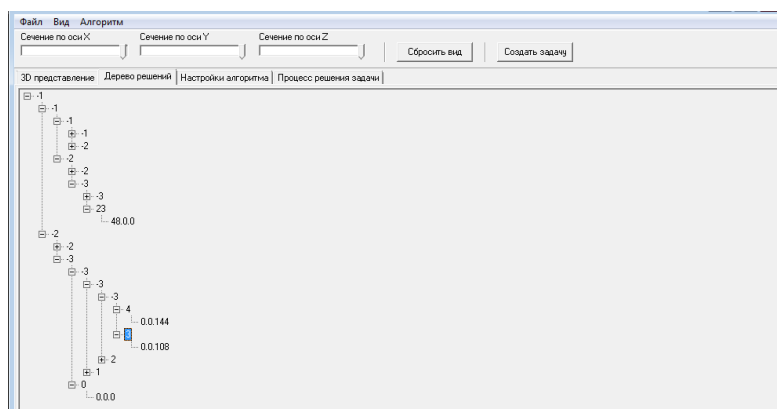


Рис. 3. Дерево решений

Разработанная программа формирует трехмерное представление оптимального размещения блоков, позволяющее оценить качество решения задачи (рис. 4). Дополнительно рассчитываются показатели эффективности упаковки, такие как плотность заполнения объема.

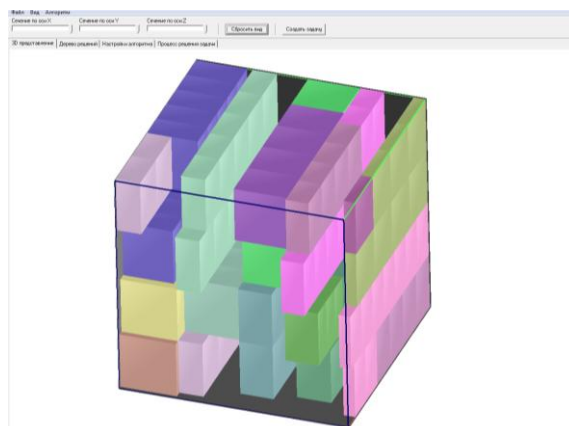


Рис. 4. Трехмерное представление результата упаковки блоков

Система предназначена для решения задач упаковки с минимизацией вычислительной сложности и адаптацией к различным условиям задачи. Эффективность разработанного подхода была проанализирована на основе результатов серии вычислительных экспериментов.

Каждый блок описывался с помощью шести параметров ($\langle k_i, p_i, l_i, w_i, h_i, v_i \rangle$), отражающих его геометрию и положение в пространстве. Критерием точности выступало отношение полноты упаковки к общему объему, а эксперименты выполнялись в 10 повторениях для каждой группы, что снижало влияние случайных факторов.

Методы расчета включали жадный поиск, локальный поиск и многоуровневый алгоритм [8]. Жадный поиск характеризуется высоким быстродействием, однако демонстрирует склонность к попаданию в локальные оптимумы. Многоуровневый алгоритм поиска представляет собой гибридный подход, сочетающий декомпозицию задачи и динамическую настройку операторов мутации и кроссинговера, что способствует адаптации к условиям задачи и достижению глобального оптимума.

$$D = \frac{\sum_{i=1}^n V_i}{V_{\text{контейнера}}} \quad (4)$$

где D – плотность упаковки; V_i – объем отдельного блока; V контейнера – объем упаковочной области

Значение D находится в пределах от 0 до 1, где 1 соответствует полностью заполненному контейнеру. Для частично заполненных контейнеров значение D рассчитывается как отношение объема упакованных блоков к объему контейнера с учетом оставшихся пустот.

Время выполнения алгоритмов оценивалось как среднее время выполнения для каждой конфигурации задач. Для минимизации случайных погрешностей выполнялось 10 запусков каждого алгоритма. Среднее время выполнения рассчитывалось по формуле (5), что позволяло объективно оценить эффективность:

$$T_{\text{ср}} = \frac{1}{k} \sum_{j=1}^k T_j \quad (5)$$

где $T_{\text{ср}}$ – среднее время выполнения; T_j – время выполнения алгоритма в j -м запуске; $j = 1$ – количество запусков.

Применение данных формул и методов было направлено на объективную оценку эффективности предложенного алгоритма, с учетом баланса между точностью и вычислительной сложностью.

Разработка многоуровневого алгоритма поиска. Исследование новых подходов к трехмерной упаковке продемонстрировало, что применение многоуровневых методов увеличивает эффективность алгоритмов оптимизации. В работе представлен адаптивный многоуровневый алгоритм поиска для оптимизации трехмерной упаковки [9].

Исследование сосредоточено на интеграции нескольких эвристических методов, что способствует улучшению качества оптимизации. Многоуровневый алгоритм включает инновационные элементы, такие как эвристика формирования начальных данных и динамическое управление генетическими операторами. Среди ключевых особенностей алгоритма: декомпозиция задачи на подзадачи для упрощения вычислений, динамическая настройка параметров для повышения адаптивности и эффективность сортировки на каждом уровне.

1. ЛПР осуществляет ввод параметров задачи, блоков, а также проводит настройку алгоритмов.

2. Генерация начальной популяции.

3. Проводится оценка созданных блоков (длина, ширина, высота, объем).

4. Осуществляется проверка блоков. Блоки, которые не удовлетворяют критериям, удаляются.

5. Сортировка и группирование блоков по группам.

6. Сортировка группы А. Блоки одинаковых размеров объединяются в отдельные подгруппы (A1, A2, A3, ..., Ai), которые помещаются в группу А. Блоки без пары помещаются в группу С.

Подгруппы группы А сортируются по объему. Подгруппы, которые содержат четное количество блоков и объем которых меньше $Y\%$ (устанавливается ЛПР) помещаются в группу В. Подгруппы, которые содержат нечетное количество блоков и объем которых меньше $Y\%$ (устанавливается ЛПР) помещаются в группу С [4].

7. Размещение блоков группы А с помощью 6-ти алгоритмов двухмерной упаковки.

8. Оценка целевой функции каждого из 6-ти решений, полученных в результате использования двухмерных алгоритмов.

9. Выбор лучшего решения, полученного в результате работы алгоритмов двухмерной упаковки.

10. Формирование группы В.

11. Формирование кубовидных моделей.

12. Размещение блоков группы В с помощью быстрого модифицированного эволюционного алгоритма.

13. Проверка на достижение одного из критериев остановки.

14. Если критерий остановки не достигнут, запускается блок эволюционной адаптации, выполняющий корректировку параметров мутации и кроссинговера для улучшения качества поиска.

15. По завершению размещения блоков группы В осуществляется формирование блоков в группу С.

16. Блоки группы С размещаются с помощью адаптированного генетического алгоритма.

17. Проверка качества полученного решения.

18. Если необходимо улучшить результат, то применяется модифицированный эволюционный алгоритм.

19. Проверка на достижение одного из критериев остановки.

20. Если критерий остановки не достигнут, то запускается блок эволюционной адаптации, в котором осуществляются настройки алгоритма.

21. После достижения одного из критериев остановки сохраняется лучшее решение.

Генетический и эволюционный алгоритмы, представленные в работе, прекратят свою работу, если одно из условий будет выполнено:

$$\Delta > \Delta Z, \quad (6)$$

где Δ – переменная, которая определяет сходимость ЦФ после каждой итерации; ΔZ – задается ЛПР.

Например, генетический алгоритм работает 300 итераций, $\Delta=10\%$ от количества итераций, а $\Delta Z=10\%$. Тогда $\Delta_x = 300 \cdot 0,1 = 30$; $300 - 30 = 270$.

Оценка ЦФ на Δ_1 270-ой итерации = 0,65.

Оценка ЦФ на Δ_2 300-ой итерации = 0,63.

$\Delta = \Delta_2 / \Delta_1 = 0,63 / 0,65 = 3\%$ на этом этапе алгоритм прекратит свою работу, так как сходимость $\Delta = 3\% < \Delta Z = 10\%$.

$$t_{\text{вып}} < T_{\text{max}}, \quad (7)$$

где $t_{\text{вып}}$ – время, которое уже затрачено на выполнение алгоритма; T_{max} – максимально допустимое время решения алгоритма.

Этот метод определяет вероятности применения кроссинговера и мутации в генетических и эволюционных алгоритмах. Основной критерий – количество итераций, при которых параметр Δ остается меньше ΔZ . Для предотвращения ухудшения результата вероятность мутации уменьшается по мере приближения к оптимальному решению [10–12]. В то же время при приближении к локальному экстремуму уменьшается вероятность кроссинговера и увеличивается применение мутации.

После активации эволюционного алгоритма осуществляется проверка выполнения условий остановки: либо достижение определенного значения ЦФ, либо превышение заранее установленного временного лимита. В случае обнаружения нового, более эффективного решения и выполнения условий остановки, фиксируется наилучший результат. В случае невыполнения критериев остановки алгоритм корректируется посредством модуля эволюционной адаптации. Поиск продолжается до выполнения одного из условий остановки.

Сравнение разработанного алгоритма с классическими методами. Одним из ключевых этапов разработки алгоритмов является их сравнение с устоявшимися методами, широко применяемыми в области оптимизации.

Проблематика упаковки и определенного размещения объектов исторически занимает центральное место в дисциплине операционных исследований. Сравнительный анализ разработанного многоуровневого алгоритма и устоявшихся методов позволяет систематически оценить их преимущества, ограничения и области применения. Сравнительный анализ разработанного многоуровневого алгоритма и традиционных методов решения:

а) Временная эффективность решения.

Традиционные методы. Как правило, используют методы экспоненциального перебора или жадные стратегии, что приводит к увеличению временных затрат при росте размерности задачи [13].

Многоуровневый алгоритм. Использует комбинацию различных стратегий с адаптивными механизмами, что, обеспечивая сокращение временных затрат и достижение решений в полиномиальное время

б) Качественная характеристика решения.

Традиционные методы. Существует вероятность застревания в локальных экстремумах, особенно при решении сложных комбинаторных задач.

Многоуровневый алгоритм. Включает механизмы, которые способствуют преодолению локальных экстремумов, повышая вероятность достижения глобально оптимального решения.

в) Область применения.

Традиционные методы. Являются результатом строгих математических разработок и могут не всегда подходить для решения разнообразных задач.

Многоуровневый алгоритм. Гибкость и масштабируемость алгоритма позволяют использовать его для решения широкого спектра задач.

г) Сложность реализации.

Традиционные методы. В большинстве случаев характеризуются относительной простотой программной реализации, используя известные и проверенные методики [14].

Многоуровневый алгоритм. Реализация алгоритма требует глубокого понимания его составляющих стратегий, что увеличивает сложность, но в то же время позволяет обеспечить высокую точность и эффективность решений

д) Адаптивные характеристики.

Традиционные методы. Зачастую работают на основе заранее установленных правил, ограниченных в возможности адаптации.

Многоуровневый алгоритм. Оснащен механизмами динамической адаптации, которые позволяют лучше учитывать изменчивые параметры задачи.

Для оценки разработанных алгоритмов была проведена серия экспериментов, включающая генетические, эволюционные и биоинспирированные алгоритмы (ГА, ЭА и БА соответственно). В ходе исследований были установлены параметры, которые характеризуются оптимальными значениями для различных операций [15–20]. Например, для оператора кроссовера (ОК) вероятность в диапазоне 60–70% продемонстрировала наилучшие результаты, а для различных операторов мутации (ОМ-1, ОМ-2, ОМ-3) – 40–50%, 45% и 35% соответственно.

Первый эксперимент был направлен на исследование зависимости значения целевой функции (ЦФ) от количества элементов при фиксированном числе итераций (200) и размере начальной популяции (25 особей). Результаты приведены рис. 5

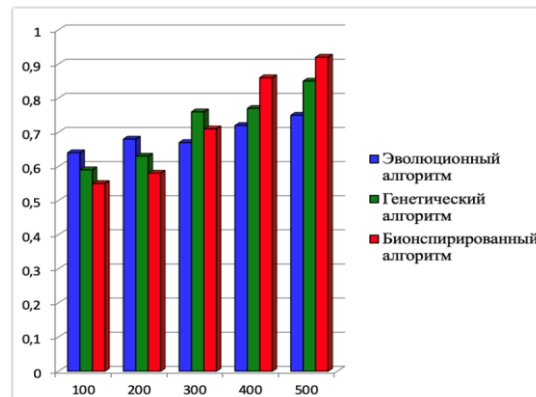


Рис. 5. График зависимости значения ЦФ от количества элементов для 200 итераций

На графике видно, что с увеличением количества блоков значение ЦФ возрастает для всех алгоритмов. Это указывает на их потенциал в решении задач с большим числом элементов. Однако даже при небольшом количестве блоков алгоритмы обеспечивают приемлемое качество решения.

Во втором эксперименте количество итераций было увеличено до 500, а размер начальной популяции – до 100 особей. Результаты представлены на рис. 6.

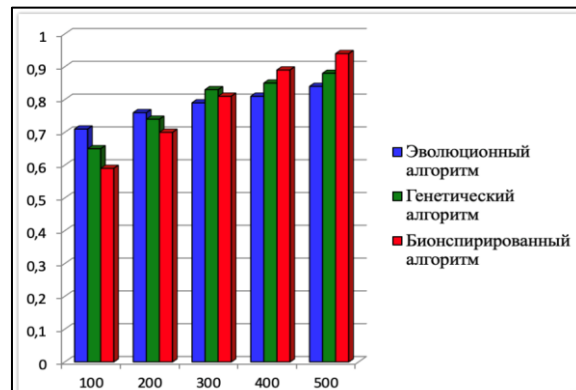


Рис. 6. График зависимости времени работы алгоритма от количества элементов для 500 итераций

С увеличением числа итераций наблюдается улучшение значений ЦФ для всех алгоритмов. Это свидетельствует о целесообразности увеличения числа итераций для повышения точности решений. Однако значительного улучшения при увеличении числа итераций не происходит, что объясняется достаточным размером начальной популяции.

Следующий эксперимент исследовал зависимость времени работы алгоритмов от числа элементов. Результаты на рис. 7.

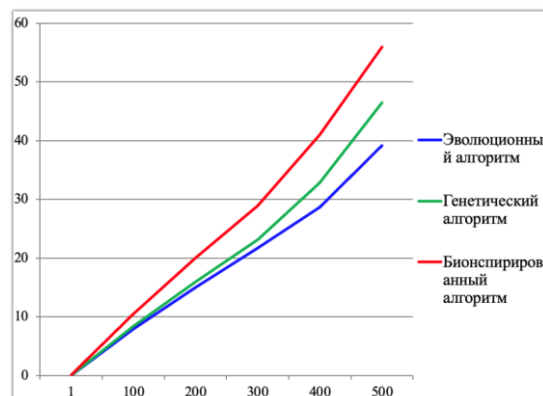


Рис. 7. График зависимости времени работы от количества элементов

Анализ графика показывает практически линейную зависимость времени работы от числа блоков. Экспериментальная временная сложность ($O(n^2)$ – $O(n^3)$) соответствует теоретическими предположениями. Наименьшее время работы показал эволюционный алгоритм, в то время как биоинспирированный алгоритм по времени близок к генетическому.

Для проверки эффективности разработанных генетических алгоритмов использовались тестовые данные из OR-Library – коллекции наборов для задач операционных исследований (OR). В табл. 2 приведено количество блоков и процент использованного пространства контейнера (%)

Таблица 2

Сравнение биоинспирированного алгоритма с аналогами

Количество блоков	Ngoi et al.	Bishoff и др.	Gehring и др.	Разработанный БА
шт.	%	%	%	%
100	62,4	62,6	62,2	63,1
200	69,1	67,3	68,1	68,9
300	75,4	74,3	76,3	78,6
400	81,8	83,9	82,8	84,9
500	88,7	89,5	91,1	92,7
Среднее значение	75,48	75,52	76,1	77,64

Сравнительный анализ разработанного многоуровневого алгоритма и традиционных методов подтвердил его преимущества в решении задач упаковки, включая увеличение плотности размещения объектов [2, 7]. Экспериментальные данные показывают, что плотность размещения объектов увеличивается в среднем на 2% по сравнению с аналогами, что эквивалентно улучшению эффективности использования пространства контейнера. Например, алгоритм может учитывать ограничения на доступное пространство, допустимую нагрузку и конфигурацию объектов. Это позволяет применять его для решения задач с различными условиями.

Преимущества и недостатки предложенного метода

Преимущества:

1. Оптимизационная эффективность. Разработанный многоуровневый алгоритм учитывает комплексность задач упаковки, предоставляя быстрый и эффективный способ решения, даже когда традиционные методы оказываются неэффективными.
2. Адаптивность к задачам. Благодаря механизмам динамической регуляции и принципам эволюционной адаптации алгоритм обладает возможностью модификации под конкретные задачи.
3. Ускорение процесса оптимизации. Многоуровневая структура в сочетании с интеграцией разнообразных методов сокращает время, необходимое для нахождения приближенных оптимальных решений.
4. Обход локальных экстремумов. Применение стратегий для преодоления локальных оптимумов способствует приближению решения к глобальному оптимальному значению.

Недостатки:

1. Техническая сложность. Интеграция многоуровневой структуры с оптимизационными методиками требует высокой квалификации и глубокого понимания алгоритмических основ, что усложняет освоение алгоритма исследователями с ограниченным опытом.
2. Затраты на вычислительные ресурсы. При определенных условиях, особенно при обработке объемных данных, алгоритм может требовать значительных вычислительных мощностей.
3. Риск переобучения, возникает при недостаточной настройке параметров и отсутствии эффективного контроля в процессе обучения модели.
4. Зависимость от исходных данных. Подготовка и предварительная обработка данных оказывают на эффективность и результаты алгоритма.

Разработанный многоуровневый алгоритм демонстрирует высокую вычислительную эффективность при решении задач оптимизации, несмотря на выявленные ограничения.

Заключение. В рамках исследования подробно изучена задача трехмерной упаковки, актуальность которой обусловлена необходимостью разработки более эффективных методик оптимизации. Анализ существующих методов, включая жадный поиск, динамическое программирование, эволюционные алгоритмы и локальный поиск, позволил выявить их ключевые характеристики и определить подходящие области применения.

На основе анализа традиционных методов был разработан многоуровневый алгоритм поиска. Данный алгоритм позволяет улучшить эффективность в сравнении с классическими методами по ряду параметров, включая скорость решения, качество решения, универсальность применения, сложность реализации и адаптивность. Разработан программный комплекс для решения задачи оптимизации трехмерной упаковки с использованием биоинспирированных алгоритмов. Проведен вычислительный эксперимент на тестовых примерах (бенчмарках). Качество упаковки, полученное, на основе разработанного комбинированного биоинспирированного алгоритма, в среднем на 7% превосходит результаты упаковки, полученные с использованием известных алгоритмов, а время решения меньше от 7% до 25%, что говорит об эффективности предложенного подхода. Проведенные серии тестов и экспериментов позволили уточнить теоретические оценки временной сложности алгоритмов упаковки. В лучшем случае временная сложность алгоритмов $O(n^2)$, в худшем случае – $O(n^3)$. Результаты анализа предоставляют пользователям возможность обоснованного выбора оптимального метода для решения задач упаковки.

Таким образом, предложенный многоуровневый алгоритм способен повысить эффективность решения задач трехмерной упаковки. Его преимущества и новизна позволяют рассматривать его как полезный инструмент для специалистов.

Исследование выполнено за счет гранта Российско-го научного фонда No 24-71-00035, <https://rscf.ru/project/24-71-00035/> в Южном федеральном университете.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Koide S., Suzuki S., Degawa S. A Palletize-Planning System for Multiple Kinds of Loads using GA Search and Traditional Search // *Intelligent Robots and Systems 95. 'Human Robot Interaction and Cooperative Robots', Proceedings.* – IEEE, 1995. – Vol. 3. – P. 510-515.
2. Курейчик В.В., Балясова Ю.В., Бова В.В. Обзор и анализ методов трехмерной упаковки при морских грузоперевозках // *Известия ЮФУ. Технические науки.* – 2024. – № 6. – С. 15-29.
3. Kang Z., Guan Y., Wang J., Chen P. Research on Genetic Algorithm Optimization with Fusion Tabu Search Strategy and Its Application in Solving Three-Dimensional Packing Problems // *Symmetry.* – 2024. – Vol. 16 (4). – P. 449.
4. Курейчик В.В., Глуценко А.Е. Многоуровневый подход для решения задачи трехмерной упаковки большой размерности // *Известия ЮФУ. Технические науки.* – 2020. – № 2 (212). – С. 6-16.
5. Курейчик В.В., Глуценко А.Е., Орлов А.Н. Гибридный подход для решения задачи 3-х мерной упаковки // *Известия ЮФУ Технические науки.* – 2016. – № 6 (179). – С. 45-53.
6. Курейчик В.В., Глуценко А.Е. Комбинированный подход для решения задачи 3-х мерной упаковки разногабаритных элементов // XII Всероссийская научная конференция молодых ученых, аспирантов и студентов, информационные технологии, системный анализ и управление (ИТСА и У-2015). – Ростов-на-Дону: Изд-во ЮФУ, 2015. – С. 75-79.
7. Курейчик В.М., Курейчик Л.В. Комплексный метод упаковки блоков // *Информатика, вычислительная техника и инженерное образование.* – 2015. – № 1 (21). – С. 17-26.
8. Шайлиев М.Б., Гарягдыев А.М. Классификация и анализ существующих алгоритмов трехмерной упаковки разногабаритных объектов в блоки // *Фундаментальные и прикладные аспекты компьютерных технологий и информационной безопасности.* – 2023. – С. 220-222.
9. Sridhar R., Chandrasekaran M., Page T. Multi-objective optimization of heterogeneous bin packing using adaptive genetic approach // *Indian Journal of Science and Technology.* – 2016. – Vol. 9, No. 48.
10. Yuan Y. et al. Three-dimensional atomic packing in amorphous solids with liquid-like structure // *Nature materials.* – 2022. – Vol. 21, No. 1. – P. 95-102.
11. Золотухин А.В., Маряшина Д.Н. Применение модифицированного генетического алгоритма как одного из эволюционных подходов к решению задачи трехмерной упаковки. – 2021.
12. Кравченко Д.Ю. и др. Биоинспирированный метод классификации распределенных вычислительных ресурсов для диспетчирования в грид-системах // *Известия ЮФУ. Технические науки.* – 2021. – № 4 (221). – С. 31-41.
13. Заруба Д.В. Компонировка блоков ЭВА на основе биоинспирированных методов поиска // *Современные компьютерные технологии.* – 2020. – С. 4-9.
14. Ling X., Osotsi M.I., Zhang W., et al. Bioinspired Materials: From Distinct Dimensional Architecture to Thermal Regulation Properties // *J Bionic Eng.* – 2023. – Vol. 20. – P. 873-899.
15. Кравченко Д.Ю. и др. Биоинспирированный метод имитационного моделирования для диспетчеризации потоков параллельных заявок в grid-системах // *Известия ЮФУ. Технические науки.* – 2020. – № 2 (212). – С. 79-88.
16. Курейчик В.В., Глуценко А.Е., Орлов А.Н. Hybrid genetic algorithm for cutting stock and packaging problems // *Proceedings of IEEE East-West Design & Test Symposium (EWDTS 2016).* Индексируется в Scopus. IEEE EWDTS 2016, Yerevan, October, 14-17, 2016. – С. 587-590.
17. Gzara F., Elhedhli S., Yildiz B.C. The pallet loading problem: Three-dimensional bin packing with practical constraints // *European Journal of Operational Research.* – 2020. – Vol. 287, No. 3. – P. 1062-1074.
18. Chen Y. et al. Recent advancements on three-dimensional electrospun nanofiber scaffolds for tissue engineering // *Advanced Fiber Materials.* – 2022. – Vol. 4, No. 5. – P. 959-986.
19. Nguyen T.H. et al. Bio-inspired approaches for smart energy management: State of the art and challenges // *Sustainability.* – 2020. – Vol. 12, No. 20. – P. 8495.
20. Jiang Y., Cao Z., Zhang J. Learning to solve 3-D bin packing problem via deep reinforcement learning and constraint programming // *IEEE transactions on cybernetics.* – 2021. – Vol. 53, No. 5. – P. 2864-2875.

REFERENCES

1. Koide S., Suzuki S., Degawa S. A Palletize-Planning System for Multiple Kinds of Loads using GA Search and Traditional Search, *Intelligent Robots and Systems 95. 'Human Robot Interaction and Cooperative Robots', Proceedings.* IEEE, 1995, Vol. 3, pp. 510-515.
2. Kureychik V.V., Balyasova Yu.V., Bova V.V. Obzor i analiz metodov trekhmernoy upakovki pri morskikh gruzoperevozkakh [Review and analysis of three-dimensional packing methods for sea freight transportation], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2024, No. 6, pp. 15-29.

3. Kang Z., Guan Y., Wang J., Chen P. Research on Genetic Algorithm Optimization with Fusion Tabu Search Strategy and Its Application in Solving Three-Dimensional Packing Problems, *Symmetry*, 2024, Vol. 16 (4), pp. 449.
4. Kureychik V.V., Glushchenko A.E. Mnogourovnevyy podkhod dlya resheniya zadachi trekhmernoy upakovki bol'shoy razmernosti [Multilevel approach for solving the problem of large-scale three-dimensional packing], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2020, No. 2 (212), pp. 6-16.
5. Kureychik V.V., Glushchenko A.E., Orlov A.N. Gibridnyy podkhod dlya resheniya zadachi 3-kh mernoy upakovki [Hybrid approach for solving the problem of 3-dimensional packing], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2016, No. 6 (179), pp. 45-53.
6. Kureychik V.V., Glushchenko A.E. Kombinirovanny podkhod dlya resheniya zadachi 3-kh mernoy upakovki raznogabaritnykh elementov [A combined approach to solving the problem of 3-D packing of different-sized elements], *XII Vserossiyskaya nauchnaya konferentsiya molodykh uchenykh, aspirantov i studentov, informatsionnye tekhnologii, sistemnyy analiz i upravlenie (ITSAiU-2015)* [XII All-Russian Scientific Conference of Young Scientists, Postgraduates, and Students, Information Technology, Systems Analysis, and Management (ITSAiU-2015)]. Rostov-on-Don: Izd-vo YuFU, 2015, pp. 75-79.
7. Kureychik V.M., Kureychik L.V. Kompleksnyy metod upakovki blokov [An Integrated Block Packing Method], *Informatika, vychislitel'naya tekhnika i inzhenernoe obrazovanie* [Computer Science, Computer Engineering, and Engineering Education], 2015, No. 1 (21). pp. 17-26.
8. Shayliev M.B., Garyagdyev A.M. Klassifikatsiya i analiz sushchestvuyushchikh algoritmov trekhmernoy upakovki raznogabaritnykh ob'ektov v bloki [Classification and analysis of existing algorithms for three-dimensional packing of different-sized objects into blocks], *Fundamental'nye i prikladnye aspekty komp'yuternykh tekhnologiy i informatsionnoy bezopasnosti* [Fundamental and Applied Aspects of Computer Technology and Information Security], 2023, pp. 220-222.
9. Sridhar R., Chandrasekaran M., Page T. Multi-objective optimization of heterogeneous bin packing using adaptive genetic approach, *Indian Journal of Science and Technology*, 2016, Vol. 9, No. 48.
10. Yuan Y. et al. Three-dimensional atomic packing in amorphous solids with liquid-like structure, *Nature materials*, 2022, Vol. 21, No. 1, pp. 95-102.
11. Zolotukhin A.V., Maryashina D.N. Primenenie modifitsirovannogo geneticheskogo algoritma kak odnogo iz evolyutsionnykh podkhodov k resheniyu zadachi trekhmernoy upakovki [Application of a modified genetic algorithm as one of the evolutionary approaches to solving the three-dimensional packing problem], 2021.
12. Kravchenko D.Yu. i dr. Bioinspirovanny metod klassifikatsii raspredelennykh vychislitel'nykh resursov dlya dispetchirovaniya v grid-sistemakh [Bioinspired method for classification of distributed computing resources for dispatching in grid systems], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 4 (221), pp. 31-41.
13. Zaruba D.V. Komponovka blokov EVA na osnove bioinspirovannykh metodov poiska [Layout of EVA blocks based on bioinspired search methods], *Sovremennye komp'yuternye tekhnologii* [Modern Computer Technologies], 2020, pp. 4-9.
14. Ling X., Osotsi M.I., Zhang W., et al. Bioinspired Materials: From Distinct Dimensional Architecture to Thermal Regulation Properties, *J Bionic Eng.*, 2023, Vol. 20, pp. 873-899.
15. Kravchenko D.Yu. i dr. Bioinspirovanny metod imitatsionnogo modelirovaniya dlya dispetcherizatsii potokov parallel'nykh zayavok v grid-sistemakh [Bioinspired simulation modeling method for dispatching parallel request flows in grid systems], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2020, No. 2 (212), pp. 79-88.
16. Kureychik V.V., Glushchenko A.E., Orlov A.N. Hybrid genetic algorithm for cutting stock and packaging problems [Hybrid genetic algorithm for cutting stock and packaging problems], *Proceedings of IEEE Ea st-West Design & Test Symposium (EWDTS 2016)* [Proceedings of IEEE Ea st-West Design & Test Symposium (EWDTS 2016)]. Indexed in Skopus. IEEE EWDTS 2016, Yerevan, October, 14-17, 2016, pp. 587-590.
17. Gzara F., Elhedhli S., Yildiz B.C. The pallet loading problem: Three-dimensional bin packing with practical constraints, *European Journal of Operational Research*, 2020, Vol. 287, No. 3, pp. 1062-1074.
18. Chen Y. et al. Recent advancements on three-dimensional electrospun nanofiber scaffolds for tissue engineering, *Advanced Fiber Materials*, 2022, Vol. 4, No. 5, pp. 959-986.
19. Nguyen T.H. et al. Bio-inspired approaches for smart energy management: State of the art and challenges, *Sustainability*, 2020, Vol. 12, No. 20, pp. 8495.
20. Jiang Y., Cao Z., Zhang J. Learning to solve 3-D bin packing problem via deep reinforcement learning and constraint programming, *IEEE transactions on cybernetics*, 2021, Vol. 53, No. 5, pp. 2864-2875.

Данильченко Владислав Иванович – Южный федеральный университет; e-mail: vdanielchenko@sfedu.ru; г. Таганрог, Россия, тел.: 88634371651; кафедра систем автоматизированного проектирования им. В.М. Курейчика; доцент.

Бова Виктория Викторовна – Южный федеральный университет; e-mail: vbova@sfedu.ru; г. Таганрог, Россия, тел.: 88634371651; кафедра систем автоматизированного проектирования им. В.М. Курейчика; доцент.

Семенова Марина Максимовна – Южный федеральный университет; e-mail: bogomolova@sfedu.ru; г. Таганрог, Россия, тел.: 88634371651; кафедра систем автоматизированного проектирования им. В.М. Курейчика; аспирант.

Игнатьева Сабина Валерьевна – Южный федеральный университет; e-mail: signateva@sfedu.ru; г. Таганрог, Россия, тел.: 88634371651; кафедра систем автоматизированного проектирования им. В.М. Курейчика; студент.

Шайлиев Марат Баталбиевич – Южный федеральный университет; e-mail: shayliev@sfedu.ru; г. Таганрог, Россия, тел.: 88634371651; кафедра систем автоматизированного проектирования им. В.М. Курейчика; аспирант.

Danilchenko Vladislav Ivanovich – Southern Federal University; e-mail: vdanielchenko@sfedu.ru; Taganrog, Russia; phone: +78634371651; V.M. Kureichik Computer-Aided Design Systems Department; associate professor.

Bova Victoria Viktorovna – Southern Federal University; e-mail: vbova@sfedu.ru; Taganrog, Russia; phone: +78634371651; V.M. Kureichik Computer-Aided Design Systems Department; associate professor.

Semenova Marina Maksimovna – Southern Federal University; e-mail: bogomolova@sfedu.ru; Taganrog, Russia; phone: +78634371651; V.M. Kureichik Computer-Aided Design Systems Department; post-graduate student.

Ignateva Sabina Valerievna – Southern Federal University; e-mail: signateva@sfedu.ru; Taganrog, Russia; phone: +78634371651; V.M. Kureichik Department of Computer-Aided Design Systems; student.

Shayliev Marat Batalbievich – Southern Federal University; Email: shayliev@sfedu.ru; Taganrog, Russia; phone: +7 88634371651; V.M. Kureichik Department of Computer-Aided Design Systems; postgraduate student.

УДК 004.93

DOI 10.18522/2311-3103-2026-1-20-33

А.Р. Недвигин, Р.М. Синецкий

АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ОЦЕНКИ ДОСТУПНОСТИ ТОВАРОВ В РАСПРЕДЕЛЕННЫХ СИСТЕМАХ ХРАНЕНИЯ НА ОСНОВЕ МЕТОДОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Представлена формализация задачи автоматизированного мониторинга наличия товаров на полках торговых объектов, и соответствия их расположения установленной планогамме с применением методов компьютерного зрения и машинного обучения. Целью исследования является разработка алгоритмического обеспечения для автоматической оценки уровня доступности товаров в распределенных системах хранения с применением методов компьютерного зрения, что позволит разрешить проблему обеспечения оптимального и необходимого ассортимента за счет контроля размещения товаров и поддержки принятия управленческих решений. Предложен технологический процесс обработки визуальных данных, включающий этапы нормализации, сегментации, локализации и классификации объектов с применением сверточных нейронных сетей, в частности YOLO и U-Net. Разработан интегральный показатель доступности, учитывающий физическую, визуальную и информационную составляющие. Сформулирована оптимизационная задача повышения уровня доступности и реализован механизм адаптивного дообучения нейронных сетей, обеспечивающий рост точности распознавания и сегментации изображений и качества аналитических рекомендаций. Предложен алгоритм повышения доступности для системы поддержки принятия решения, основанный на построении маршрута мерчендайзера, который учитывает приоритетность товаров и минимизирует временные затраты. Задача построения

маршрута мерчандайзера сведена к обобщенной TSP-задаче с весами приоритетов. Предложены и описаны методы оценки и повышения уровня доступности товаров. На основе предложенных методов и алгоритмов разработан программный комплекс для оценки и повышения показателя доступности товаров. Результаты исследования подтверждают эффективность предложенных решений - средняя точность распознавания составила 95,8%, интегральный показатель доступности $A = 0.93$. Практическая значимость работы заключается в обеспечении алгоритмической основы для создания интеллектуальных систем мониторинга доступности товарных запасов, способствующих повышению эффективности управления бизнес-процессами.

Мониторинг товаров; системы хранения; компьютерное зрение; сверточные нейронные сети; машинное обучение.

A.R. Nedvigin, R.M. Sinetsky

METHODOLOGICAL SUPPORT FOR ASSESSING THE AVAILABILITY OF GOODS IN DISTRIBUTED STORAGE BASED ON COMPUTER VISION METHODS

This paper presents a formalization of the problem of automated monitoring of product availability on retail shelves and compliance with the prescribed planogram, leveraging computer vision and machine learning techniques. The aim of this research is to develop algorithmic solutions for the automatic assessment of product availability in distributed retail environments using computer vision methods, thereby addressing the challenge of maintaining optimal and necessary product assortments through continuous shelf monitoring and supporting data-driven managerial decision-making. A technological pipeline for visual data processing is proposed, comprising the stages of image normalization, segmentation, object localization, and classification, implemented with convolutional neural networks—specifically YOLO and U-Net architectures. An integrated product availability metric is introduced, which jointly accounts for physical, visual, and informational dimensions of availability. An optimization problem aimed at improving overall availability is formulated, and an adaptive neural network fine-tuning mechanism is implemented to enhance the accuracy of image recognition and segmentation, as well as the quality of analytical recommendations. Furthermore, an availability-improvement algorithm is proposed for a decision support system, based on the construction of an optimized merchandiser routing plan that prioritizes products and minimizes time expenditures. This routing problem is reduced to a generalized Traveling Salesman Problem (TSP) with priority-based weights. Methods for evaluating and enhancing product availability are proposed and described in detail. Based on the developed approaches and algorithms, a software system for monitoring and improving product availability has been implemented. Experimental results confirm the effectiveness of the proposed solutions: the average recognition accuracy reached 95.8%, and the integrated availability score achieved $A = 0.93$. The practical significance of this work lies in establishing an algorithmic foundation for intelligent shelf-monitoring systems that enable more efficient management of retail operations and inventory processes.

Product monitoring; storage systems; computer vision; convolutional neural networks; machine learning.

Введение. Актуальность исследования обусловлена необходимостью повышения эффективности управления товарными запасами и обеспечением высокой доступности товаров в условиях динамичного рынка розничной торговли. Несмотря на развитие технологий компьютерного зрения и аналитических систем, остается нерешенной проблема формализованной количественной оценки уровня доступности и разработки методов ее оптимизации. Задачу автоматизации мониторинга уровня доступности товаров можно рассматривать применительно к сфере розничной торговли как комплексную проблему обеспечения оптимального ассортимента и улучшения покупательского опыта за счет контроля размещения товара на полках и обеспечения принятия управленческих решения в режиме реального времени. Ее можно решить при помощи алгоритмов компьютерного зрения и машинного обучения, объединенных в информационно-аналитической системе, выполняющей следующие функции: обнаружение товаров на изображениях, классификацию по категориям, сверку фактического размещения с планограммой, генерацию рекомендаций, формирование аналитических отчетов.

Такая система может быть создана с применением методов компьютерного зрения (Computer Vision, CV) на основе глубокого обучения (Deep Learning, DL) [1]. DL представляет собой подраздел машинного обучения (Machine Learning, ML), которое, в свою

очередь, является областью искусственного интеллекта (рис. 1). *ML* улучшает производительность алгоритмов на заданных задачах путем анализа данных и выявления закономерностей без прямого программирования конкретных решений. *DL* автоматически извлекает высокоуровневые признаки из сырых данных, что делает его эффективным для решения сложных задач, например в области распознавания образов. Одним из наиболее успешных решений в контексте обработки визуальной информации являются сверточные нейронные сети (*Convolutional Neural Networks, CNN*). Это подкласс глубоких нейронных сетей, разработанных для эффективной обработки изображений, предоставляющий переход от низкоуровневых признаков (контур или цвет) к высокому уровню абстракции – объекту в целом.

CV фокусируется на анализе и интерпретации содержимого изображений с использованием *DL* (в частности, *CNN*) для обнаружения объектов, сегментации и классификации. Иерархию и взаимосвязь концепций можно описать так: *ML* – общая парадигма системы, *DL* – специализированный инструмент *ML* для работы с неструктурированными данными (изображения, текст, звук), *CV* – прикладная область, в которой *DL* используется для решения задач анализа визуальной информации.

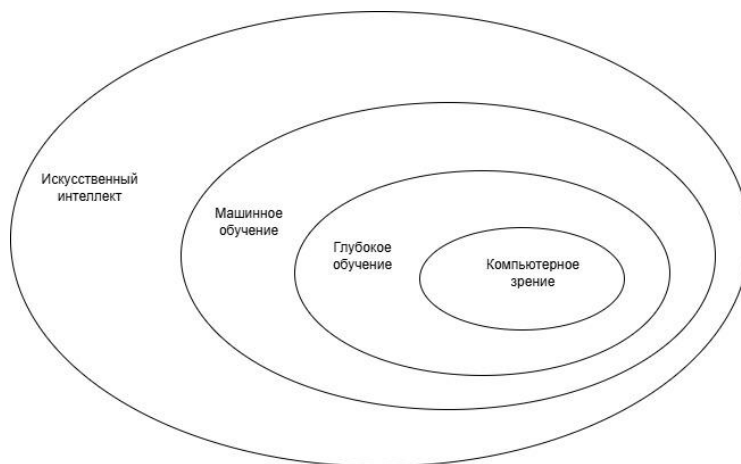


Рис. 1. Схема связи между глубоким обучением и компьютерным зрением

Обзор современных научных исследований [2–5] показывает востребованность и актуальность проводимых работ в области автоматизации мониторинга состояния торгового пространства в целом и автоматизации задач оценки доступности товара, в частности. При этом, остается недостаточно изученной и освещенной комплексная задача оптимизации, направленной на повышение уровня доступности товаров. В связи с чем необходимо разработать методику, как совокупность методов и алгоритмов, а также способов их применения для оценки уровня доступности объектов в распределенных системах с применением методов компьютерного зрения, а также сформулировать оптимизационную задачу повышения уровня доступности, предоставив основу для реализации систем для автоматизации решения задач мониторинга состояния товарной представленности и соответствия установленной выкладке.

Алгоритмы обработки визуальных данных. В качестве входных данных принимаются изображения торгового оборудования размером $n \times m$, полученные при помощи камеры смартфона:

$$I_{input} = \begin{bmatrix} I_{(0,0)} & & \\ & \ddots & \\ & & I_{(n,m)} \end{bmatrix},$$

где n – число строк, m – число столбцов.

Изображения могут содержать как непосредственно сами товары на полках, так и элементы фона – стеллажи и прочие объекты. Выходные данные – отчеты о наличии либо отсутствии товаров в установленном количестве, список выявленных отклонений от программы и рекомендации по их устранению. Устанавливаются ограничения системы: точность распознавания отдельных товаров на изображении не менее 95%, оптимизация временных затрат на обработку одного изображения до 1 секунды и возможность масштабирования системы для применения в торговых объектах вне зависимости от их размера.

Предварительная обработка изображений. Рассмотрим этапы предобработки:

♦ Нормализация изображений, установка стандартизированного размера 640×640 пикселей для унификации входящих данных, чтобы уменьшить вычислительные затраты и облегчить работу модели:

$$I_{normalized}(x, y) = \frac{I_{input}(x, y) - \mu}{\sigma}, \quad x = 0, \dots, m - 1, \quad y = 0, \dots, n - 1,$$

где I_{input} – исходное изображение размером $m \times n$ ($m = 640$, $n = 640$); μ – среднее значение пикселей всего изображения I_{input} , σ – стандартное отклонение значений пикселей изображения I_{input} .

♦ Устранение шумов с применением фильтров (Гаусса и медианного) для удаления артефактов с изображений [6].

Медианный фильтр является локальным оператором, в котором значения пикселя с координатами (x, y) отфильтрованного изображения I_{med} вычисляется как медиана значений пикселей исходного изображения в области размером $K \times K$ ($K = 2k + 1$, $k \geq 0$) вокруг пикселя исходного изображения $I_{normalized}$ с центром в точке (x, y) :

$$I_{med}(x, y) = \text{median}(\{I_{normalized}(x + i, y + j)\}), \quad i = -k, \dots, k, \quad j = -k, \dots, k.$$

Фильтр Гаусса реализуется как локальный оператор свертки, ядро которого размером $K = 2k + 1$, $k \geq 0$ вычисляется как двумерная функция Гаусса:

$$G(i, j) = \frac{1}{2\pi s^2} \exp\left(-\frac{i^2 + j^2}{2s^2}\right), \quad i = -k, \dots, k, \quad j = -k, \dots, k,$$

где s – параметр фильтрации, задающий интенсивность сглаживания изображения.

Значение пикселя с координатами (x, y) отфильтрованного изображения I_{blur} определяется как свертка с исходным изображением I_{med} :

$$I_{blur}(x, y) = \sum_{i, j=-k}^{+k} I_{med}(x + i, y + j) \cdot G(i, j), \quad x = 0, \dots, m - 1, \quad y = 0, \dots, n - 1.$$

♦ Усиление контрастности в целях улучшения различимости объектов на изображении с использованием нелинейной коррекции изображения методом гистограммной эквализации с приведением гистограммы распределения значений пикселей изображения к нормальному закону [6].

♦ Аугментация данных (повороты, сдвиги, изменения яркости) для увеличения разнообразия обучающих данных и повышения робастности модели и предотвращения возможности переобучения [7].

Сегментация объектов. Методы сегментации могут включать классическую сегментацию на основе пороговых значений, либо глубокие нейронные сети для сегментации, например, *U-Net* [8], которые позволяют выделить более сложные структуры. На этом этапе выполняется выделение областей интереса, которые соответствуют полкам либо секциям торгового зала с помощью *CNN*. Цель – ограничить область поиска товаров на изображении, что позволяет значительно ускорить последующую обработку и улучшить производительность системы.

Локализация объектов. Для обнаружения и идентификации объектов на изображениях применяется сверточная нейронная сеть *YOLO (You Only Look Once)* [9], которая анализирует изображение целиком и предсказывает классы и ограничивающие рамки

объектов за один проход. *YOLOv9* имеет модульную структуру, состоящую из слоев свертки, блока *Neck* (для обработки и объединения признаков) и блока *Head* (генерирует окончательные предсказания классов и координат объектов). Также предусмотрены слои обработки данных и слой постобработки для удаления дублирующих предсказаний. Математически алгоритм *YOLO* сводится к задаче оптимизации координат объектов и вероятностей их классов:

$$p(x, y) = \sigma_{act}(\theta_i * x + b_i),$$

где x, y – координаты центра объекта на изображении, θ_i, b – веса и смещения для каждого объекта в классе, σ_{act} – функция активации для предсказания вероятности присутствия объектов.

Процесс предсказания ограничительных рамок в алгоритме YOLO. Процесс предсказания ограничительных рамок (*bounding box*) является одним из ключевых этапов работы алгоритма [10]. *YOLO* разбивает изображение на равномерную сетку размером $S \times S$. Каждая ячейка предсказывает рамки с параметрами:

- ♦ Координаты центра рамки (x, y) – нормализованные значения относительно границ ячейки.
- ♦ Размеры рамки (w, h) – ширина и высота, выраженные в долях от размеров исходного изображения.
- ♦ Вероятность наличия объекта внутри рамки (P_{obj}) – оценка достоверности предсказания.
- ♦ Вероятности класса ($P(C_1|obj), P(C_2|obj), \dots$) – условные вероятности принадлежности объекта к C – классу, где C – идентификатор товара.

Для повышения точности предсказания рамок объектов используются якорные рамки (*anchor box*) – заданные эталонные формы, которые адаптируются к типичным пропорциям объектов в выборке (рис. 2). Координаты рамки вычисляются следующим образом:

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_w = p_w e^{t_w} \\ b_h = p_h e^{t_h} \end{cases},$$

где (c_x, c_y) – координаты верхнего левого угла ячейки, (p_w, p_h) – размеры *anchor box*, t_x, t_y, t_w, t_h – выходные данные нейросети, σ – сигмоидальная функция, ограничивающая смещение центра рамки пределами ячейки.

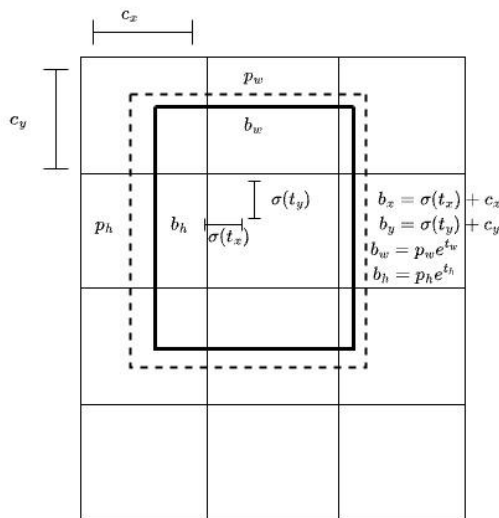


Рис. 2. Предсказание местоположения с использованием *anchor boxes*

Кроме того, для каждой ячейки рассчитываются вероятности P_{obj} :

$$P_{obj} = C(obj) * IoU(Pred, Truth).$$

Для обучения модели используется составная функция потерь, которая включает компоненты, отвечающие за точность предсказания рамок. Основной метрикой является среднеквадратичная ошибка (MSE) между предсказанными и истинными значениями (x, y, w, h) .

$$\lambda_{MSE} = \sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2],$$

где N – количество объектов на изображении.

Для фильтрации дублирующих рамок применяется алгоритм *Non-Maximum Suppression (NMS)* [11]. Происходит отсев рамок с $P_{obj} < \theta$, где θ – порог, подобранный эмпирическим путем. Для каждой вычисляется IoU [12], и если показатель перекрывающихся рамок превышает заданный порог, то рамка с более низкой оценкой удаляется.

Классификация товаров. Классификация товаров выполняется с помощью *CNN*, обученной на заранее размеченных данных и относится к задаче многоклассовой классификации [13], где товар относится к одному из классов, например, категории или бренду. Этапы классификации можно описать так:

- ◆ Извлечение признаков при помощи сверточных слоев, выполняющими фильтрацию изображения для выделения ключевых признаков: контуров, текстур и форм.
- ◆ Классификация с использованием полносвязных слоев, которые вычисляют вероятность принадлежности товара к одному из конкретных классов. Математически процесс классификации можно представить следующим образом:

$$y = f(W_x + b),$$

где W – веса, обученные в процессе тренировки сети, b – смещение, а x – это извлеченные признаки изображения товара, f – функция активации, в частности, *softmax* для многоклассовой классификации.

Обнаружение пустых мест. Выявление пустых участков на полках осуществляется на основе пространственного анализа. Используются морфологические операции эрозии и дилатации, позволяющие выделить области, не содержащие данных [14]:

Эрозию можно представить как операцию уменьшения объекта на изображении:

$$E(x, y) = \min (I(x, y) \oplus S),$$

где $\oplus$ – операция морфологической свертки, а S – структурный элемент

Дилатацию же понимают как операцию, которая расширяет объекты:

$$D(x, y) = \min (I(x, y) \odot S),$$

где $\odot$ – операция морфологической дилатации.

Дополнительно проводится анализ структур для разделения фона и объектов, что позволяет локализовать незаполненные участки [15].

Постобработка и фильтрация. После обнаружения товаров важно провести постобработку: удаление ложных срабатываний при помощи фильтров и удаление артефактов (пересекающихся объектов). Для разделения соприкасающихся товаров и устранения искажений из-за частичного перекрытия применяются методы глубокого обучения для сегментации, в частности, использование модели *U-Net* для точного выделения контуров объектов.

Методы сверки фактического размещения с планограммой. Планограмма – графическое отображение расположения товаров на полках, преобразуемое в цифровой формат. Рассмотрим три возможных представления: матрица, таблица и геометрическое представление [16].

Матричное представление разбивает полку на сетку, где каждый элемент сетки соответствует определенному месту на полке с данными о товаре.

$$P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{m1} & P_{m2} & \dots & P_{mn} \end{bmatrix},$$

где P_{ij} – это данные о товаре j -го типа на i -й полке.

Преимущества: точное позиционирование и простота анализа. Из недостатков – большие затраты по памяти при увеличении множества товаров.

Табличное представление использует базы данных, присваивая товарам координаты и параметры:

$$T = \{(ID, Category, Quantity, x, y)\},$$

где ID – это уникальный идентификатор товара, $Category$ – категория товара, $Quantity$ – количество товара, которое должно быть на полке, x, y – координаты товара на полке в двумерной системе.

Табличное представление обладает большей гибкостью и позволяет хранить разнобразную информацию о товаре и обладает возможностями для масштабирования. Недостаток: нужно явно указывать координаты для каждого товара, что усложняет сравнение с фактическим расположением.

Геометрическое отображает планограмму как набор прямоугольных областей, где каждый объект имеет определенные координаты и размеры:

$$R = \{(x_{min}, y_{min}, x_{max}, y_{max}, ID)\},$$

где (x_{min}, y_{min}) – координаты верхнего левого угла прямоугольника, (x_{max}, y_{max}) – координаты нижнего правого угла прямоугольника, ID – уникальный идентификатор товара, размещенного в этой области.

Геометрическое представление является наглядным и легко интерпретируемым. Для сверки расположения товаров с планограммой удобно использовать IoU (*Intersection over Union*). При этом требуется точное указание размеров и положения каждого товара, что увеличивает трудоемкость при работе с большими объемами данных.

Сопоставление координат и вычисление IoU для каждого товара. Для сверки фактического размещения с планограммой используется метрика IoU . Она определяет степень перекрытия между фактическими координатами товаров с предполагаемыми согласно планограмме:

$$IoU = \frac{S_{intersection}}{S_{union}},$$

где $S_{intersection}$ – это площадь, где пересекаются фактическое и предполагаемое расположение товара, а S_{union} – это площадь, занимаемая обеими объединенными областями.

Определяются случаи, в которых IoU меньше установленного порога (в конкретном случае 0,8). В таком случае фиксируется отклонение, так как считается, что товар расположен вне установленной планограммы, либо отсутствует. Схема процесса сверки размещения товара с планограммой представлена на рис. 3.

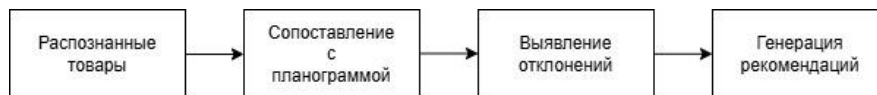


Рис. 3. Схема процесса сверки фактического размещения товара с планограммой

Методы оценки уровня доступности объектов. В общем виде цель задачи по оценке уровня доступности товаров на полках розничных магазинов можно сформулировать как определение степени фактического наличия товаров их плановому состоянию, заданному планограммой. Система оценки доступности опирается на концепцию обратной связи, при которой измеренные параметры торгового пространства сравниваются с целевыми значениями и формируют сигнал ошибки, используемый для принятия управленческих решений.

Для оценки уровня доступности введем интегральный показатель A , учитывающий три составляющих:

1. Физическая доступность (наличие товара на полке).

$$A_f = \frac{N_{present}}{N_{total}},$$

где $N_{present}$ – количество позиций, фактически присутствующих на полке, N_{total} – общее количество позиций, подлежащих проверке.

2. Визуальная доступность (доля видимой площади упаковок относительно эталонной)

$$A_v = 1 - \frac{S_{empty}}{S_{total}},$$

где S_{empty} – площадь пустого пространства, где упаковки отсутствуют, но должны быть по нормативу, S_{total} – общая площадь для размещения товара по плануграмме.

3. Информационная доступность (соответствие фактического размещения плануграмме).

$$A_i = \frac{N_{correct}}{N_{checked}},$$

где $N_{correct}$ – количество проверенных позиций, корректно размещенных в соответствии с плануграммой, $N_{checked}$ – общее количество проверенных позиций, для которых проводилось сравнение с плануграммой.

Интегральный показатель определяется как:

$$A = \omega_1 * A_f + \omega_2 * A_v + \omega_3 * A_i,$$

где $\omega_1 + \omega_2 + \omega_3 = 1$ – весовые коэффициенты, определяемые методом экспертных оценок или, иначе, методом аналитической иерархии – *AHP* (*Analytic Hierarchy Process*) [17].

При полученном значении A ниже заданного порога (определяется эмпирически, например 0,8) потребуются корректирующие воздействия для повышения доступности товаров.

Генерация рекомендаций и алгоритм поддержки принятия решений. На основе результатов оценки уровня доступности генерируется список действий для устранения выявленных отклонений, например переместить товар на полку согласно плануграмме или добавить недостающий товар для заполнения полки. Но так список рекомендаций будет сформирован без достаточной гибкости, необходимой для систем принятия решений, так как не учитывает приоритеты пополнения товаров. Выделим такие факторы, как товары первой необходимости (индикаторы), влияющие на выбор покупателя, акционные товары, наличие которых определяется договорными обязательствами, высокомаржинальные товары (продукция собственной торговой марки). Соответственно, для уточненных рекомендаций необходимо внести в алгоритм поправочные коэффициенты на основе статистики продаж. В рамках данной работы предложено упрощение модели с использованием базовых показателей: s – средняя оборачиваемость товара за день (шт. /день), p – цена товара (руб.), u – срочность ($0 \dots 100$), где 100 – максимальная срочность;

Для каждого обнаруженного класса i введем коэффициент приоритета P_i :

$$P_i = w_s * \frac{s_i}{\max_j s_j} + w_p * \frac{p_i}{\max_j p_j} + w_u * \frac{u_i}{100},$$

где w_s, w_p, w_u – весовые коэффициенты (сумма которых равна 1), нулевое значение является собой отсутствие отклонений, а единица – максимальное отклонение, требующее приоритетного устранения.

Для определения весовых коэффициентов используется метод *AHP*. Пример расчета: позиция A имеет $s_A = 0,8$, что эквивалентно 80% от максимального объема продаж, $p_A = 0,5$ (половина от максимальной цены), $u_A = 1$ (высокая срочность). При $w_s = 0,3$, $w_p = 0,2$, $w_u = 0,5$ мы получим $P_A = 0,84$.

Построение маршрута мерчендайзера. Пусть мерчендайзер получил список товарных позиций (*stock keeping unit, SKU*) $\{i_1, i_2, \dots, i_k\}$ с приоритетами $\{P_1, P_2, \dots, P_k\}$. Построение маршрута мерчендайзера сводится к задаче коммивояжера с весами приоритетов (*travelling salesman problem, TSP*) [18]. Целью является минимизация общего расстояния и времени (стоимости) пути с учетом приоритетов. Задача формулируется как

обобщенная *TSP* с весами приоритетов, где граф $G = (V, E)$, где V – набор точек, представляющих собой товарные позиции на полке, E – ребра с весами $d(v_i, v_j)$. Оптимальный маршрут π определяется как минимум от обобщенной длины:

$$L(\pi) = \sum_{t=1}^{k-1} d(v_{\pi(t)}, v_{\pi(t+1)}) + \alpha \sum_{t=1}^k (t-1)(1 - P_{\pi(t)}),$$

где первый член – расстояние пути, а второй член – штрафной, который способствует стимулированию посещения узлов с высоким P .

Используется жадный алгоритм с оценочной функцией: $Score(j) = \lambda d(v_{current}, v_j) - (1 - \lambda)P_j$, где $\lambda \in [0, 1]$ – параметр между сокращением расстояния и учетом приоритета. Ниже пример расчета маршрута мерчендайзера. Данные для расчета примера см. в табл. 1.

Таблица 1

Исходные данные

SKU	Оборачиваемость s_i (шт./день)	Стоимость p_i (руб)	Срочность u_i (%)	X(м)	Y(м)
Позиция А	120	100	100	20	8
Позиция В	80	150	40	5	15
Позиция С	100	120	60	8	2
Позиция D	50	80	20	3	3
Позиция Е	150	20	100	7	18

- ♦ Оборачиваемость – среднее число проданных единиц товара в день.
- ♦ Стоимость – средняя цена товара.
- ♦ Срочность – показывает отличие наличие товара отличается от заданного в программе. Примем, что А и Е полностью отсутствуют, т.е. имеют отклонение равное 100%, актуальный остаток В составляет 60%, т.е. отклонение 40% и т.д.
- ♦ Координаты X, Y – положение товаров на плане торгового зала (метры от стартовой точки (0,0)). Каждый показатель нормируется к диапазону [0,1] (табл. 2) по максимальному значению в столбце: $\hat{s}_i = \frac{s_i}{\max(s)}$, $\hat{p}_i = \frac{p_i}{\max(p)}$, $\hat{u}_i = \frac{u_i}{100}$.

Таблица 2

Данные после нормирования

SKU	$\hat{s}_i$	$\hat{p}_i$	$\hat{u}_i$
Позиция А	0.8	0.67	1.0
Позиция В	0.53	1.0	0.4
Позиция С	0.67	0.8	0.6
Позиция D	0.33	0.53	0.2
Позиция Е	1.0	0.13	1.0

Для расчета приоритета P_i используем весовые коэффициенты $w_s = 0,3$, $w_p = 0,2$, $w_u = 0,5$. Таким образом, получаем: $P_i = w_s \hat{s}_i + w_p \hat{p}_i + w_u \hat{u}_i$

Для построения принятия решений очередности выполнения заданий используем алгоритмическую парадигму жадной эвристики [19]. Пусть $\lambda = 0,5$. На каждом шаге среди непосещенных узлов выбирается тот, у которого минимально значение $Score(j) = \lambda d(current, j) - (1 - \lambda)P_j$, где d – евклидово расстояние от текущей позиции. Таким образом, итоговый маршрут выглядит следующим образом: начало маршрута:

(0,0) – стартовая точка, вход в торговый зал. Последовательность отработки заданий: $D(3,3) \rightarrow C(8,2) \rightarrow A(20,8) \rightarrow B(5,15) \rightarrow E(7,18) \rightarrow (0,0)$. График рассчитанного маршрута см. на рис. 4.

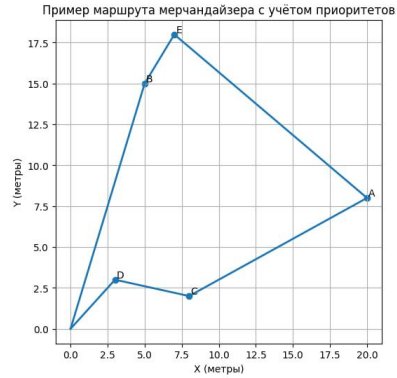


Рис. 4. Пример маршрута мерчендайзера с учетом приоритетов

Механизм обратной связи для дообучения. Для минимизации повторяющихся ошибок и сокращения предусмотрим алгоритм цикла обратной связи. В случае, если сотрудник отмечает задание как ошибочное («Отклонить – Ложное срабатывание»), то система отмечает запись соответствующим флагом в *DETECTIONS*, а в таблицу *LOGS* добавляется запись, *action = false_positive* для соответствующего *detection_id*. Счетчик ложных срабатываний *false_positive_count* в *DETECTIONS* увеличивается при каждом отклонении для класса *i* в рамках одной задачи. При достижении значения «3», формируется запрос на перестройку тренировочной выборки. Ложные примеры добавляются в пакет для дообучения.

Система формирует выборку верно положительных детекций и ложно положительных сигналов в пропорции 4:1. При достижении ≥ 100 ложно положительных примеров запускается инкрементальный процесс дообучения без полного переобучения модели. Диаграмма процесса показана на рис. 5.

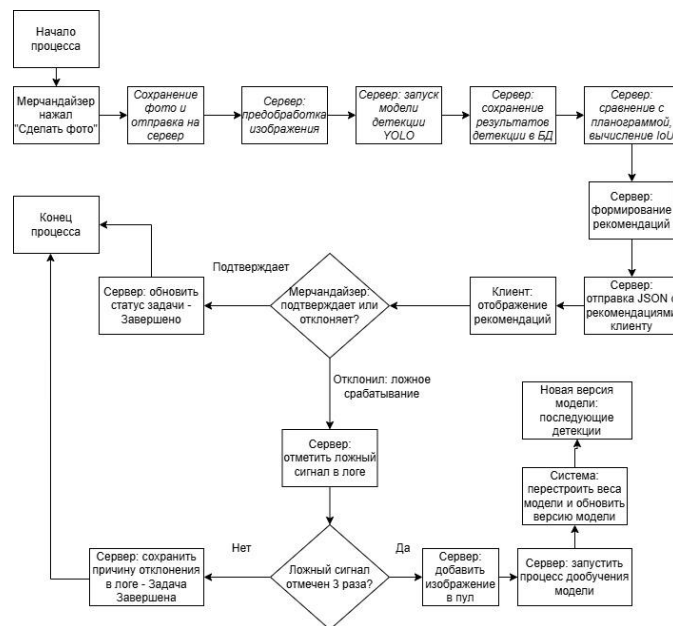


Рис. 5. Диаграмма процесса потока работ до возможного дообучения

Генерация аналитических отчетов. Основные типы отчетов включают в себя:

- ♦ Отчет о доступности товаров, который показывает процент заполнения полок и выделяет пустые места, чтобы оптимизировать запасы.

$$P_{empty} = \frac{S_{empty}}{S_{total}} \times 100,$$

где S_{empty} – площадь пустых мест на полке, S_{total} – общая площадь полки.

- ♦ Отчет о соблюдении планограммы, в котором указываются выявленные отклонения и несоответствия в выкладке товаров.

Оценка производительности системы. Для эффективной интеграции системы в торговую среду, необходимо произвести оценку ее производительности. Основные метрики [20]:

- ♦ Точность распознавания (*Precision*). определяет долю верно классифицированных примеров:

$$P = \frac{TP}{TP+FP},$$

где TP (*True Positives*) – количество истинно-положительных результатов; FP (*False Positives*) – количество ложных положительных результатов.

- ♦ Полнота (*Recall*) – доля корректно классифицированных положительных примеров:

$$P = \frac{TP}{TP+FN},$$

где FN (*False Negatives*) – количество ложных отрицательных результатов.

- ♦ Среднее время обработки одного изображения.

- ♦ Скорость обработки в реальном времени.

Задача повышения уровня доступности. Повышение уровня доступности товаров рассматривается как комплексная задача оптимизации, которая включает в себя диагностический, аналитический, оптимизационный и корректирующий этапы.

1. Диагностический этап. Направлен на фиксацию текущего состояния системы и определение факторов, влияющих на снижение доступности. Включает в себя сбор данных, анализ изображения, сверку с эталонной схемой, расчет частных показателей доступности (физической, визуальной, информационной).

2. Аналитический этап. Для каждого найденного класса объектов определяется коэффициент приоритета.

3. Оптимизационный этап. На основе выявленных факторов строится модель оптимизации доступности. Алгоритм определяет оптимальную последовательность действий, минимизирующую совокупные затраты.

4. Корректирующий этап. Рассматривается как формирование управленческих воздействий с созданием перечня корректирующих мероприятий. В их числе перемещение или добавление товаров согласно эталонной планограмме, корректировка приоритетов, а также формирование маршрута обхода с учетом приоритетов с обеспечением минимизации пути и времени и максимизации прироста A .

5. Этап адаптивного дообучения.

Полный цикл процесса повышения уровня доступности можно представить как замкнутую систему управления (рис. 6).



Рис. 6. Схема процесса повышения уровня доступности объектов

На каждом цикле функция управления $U(t)$ определяется как:

$$U(t) = f(E_x, P_i, \Delta A_t),$$

где P_i – приоритет позиции, E_x – эталонное состояние, $\Delta A_t = A_{t+1} - A_t$, что отражает прирост доступности по результатам цикла. При этом, эффективность оценивается сравнением значений A_t и A_{t+1} , рост показателя A свидетельствует о корректности механизмов диагностики и оптимизации.

Заключение. Разработаны алгоритмы оценки уровня доступности товаров в распределенных системах хранения с применением методов компьютерного зрения. Предложен интегральный показатель доступности, учитывающий физическую, визуальную и информационную составляющие. Сформулирована оптимизационная задача повышения доступности и реализован механизм адаптивного дообучения, обеспечивающий рост показателя доступности. На основе предложенных методов и алгоритмов разработан программный комплекс [21] для оценки и повышения показателя доступности товаров. Средняя точность распознавания объектов при помощи разработанного программного комплекса составила 95,8%, показатель доступности $A = 0,93$. На основании проделанной работы можно утверждать, что предложенная совокупность методов обнаружения идентификации товаров, извлечения признаков, классификации и формирования аналитических отчетов за счет комбинирования методов глубокого и машинного обучения является собой эффективный и надежный способ мониторинга торгового пространства.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Tiwari Tanya & Tiwari Tanuj & Tiwari Sanjay.* How Artificial Intelligence, Machine Learning and Deep Learning are Radically Different? // *International Journal of Advanced Research in Computer Science and Software Engineering.* – 2018. – 8(2):1. – 10.23956/ijarcsse.v8i2.569.
2. *Mehwish Saqlain, Saddaf Rubab, Malik M. Khan, Nouman Ali, Shahzeb Ali.* Hybrid Approach for Shelf Monitoring and Planogram Compliance (Hyb-SMPC) in Retails Using Deep Learning and Computer Vision // *Mathematical Problems in Engineering.* – 2022. – No. 1. – <https://doi.org/10.1155/2022/4916818>.
3. *Jorge Muñoz, Alonso Sanchez, Guillermo Kemper.* End-to-end solution for automatic beverage stock detection in supermarkets based on image processing and convolutional neural networks // *International Journal of Cognitive Computing in Engineering.* – 2024. – No. 5. – P. 453-474. – <https://doi.org/10.1016/j.ijcce.2024.09.001>.
4. *Wei Yuchen, Tran Son, Xu Shuxiang, Kang Byeong, Springer Matthew.* Deep Learning for Retail Product Recognition: Challenges and Techniques // *Computational Intelligence and Neuroscience.* – 2020. – 8875910. – P. 1-23. – <https://doi.org/10.1155/2020/8875910>.
5. *Gothai E. et al.* Design features of grocery product recognition using deep learning // *Intelligent Automation and Soft Computing.* – 2022. – Vol. 34, No. 2. – P. 1231-1246.
6. *Рейнхард Клетте.* Компьютерное зрение. Теория и алгоритмы: пер. с англ. А.А. Слинкин. – М.: ДМК Пресс, 2019. – 506 с.
7. *Yolo Data Augmentation.* – 2025. – Режим доступа: <https://docs.ultralytics.com/guides/yolo-data-augmentation> (дата обращения: 02.05.2025).
8. *Surono Sugiyarto & Rivaldi Muhammad & Arrova Dewi Deshinta & Irsalinda Nursyiva.* New Approach to Image Segmentation: U-Net Convolutional Network for Multiresolution CT Image Lung Segmentation // *Emerging Science Journal.* – 2023. – 7. – P. 498-506. – 10.28991/ESJ-2023-07-02-014.
9. *YOLOv9 Architecture Explained.* – Режим доступа: <https://stunningvisionai.com/article/yolov9-architecture> (дата обращения: 02.05.2025).
10. *Desai Miral & Mewada Hiren & Pires Ivan & Roy Sparsh.* Evaluating the Performance of the YOLO Object Detection Framework on COCO Dataset and Real-World Scenarios // *Procedia Computer Science.* – 2024. – 251. – P. 157-163. – 10.1016/j.procs.2024.11.096.
11. *Wang Decheng & Chen Xiangning & Yi Hui & Zhao Feng.* Improvement of Non-Maximum Suppression in RGB-D Object Detection // *IEEE Access.* – 2019. – Vol. 7. – 10.1109/ACCESS.2019.2945834.
12. *Бондаренко В.А., Гессен П.А., Павлова В.А., Созинова М.В., Туников В.А.* Алгоритм обнаружения объектов для оптико-электронных систем с обучением в реальном времени // *Известия ЮФУ. Технические науки.* – 2021. – № 1 (218).
13. *Бринк Х., Ричардс Д., Февеолф М.* Машинное обучение. – СПб.: Питер, 2017. – 336 с.
14. *Said K & Jambek Asral.* Analysis of Image Processing Using Morphological Erosion and Dilation // *Journal of Physics: Conference Series.* – 2021. – 2071(1):012033. – 10.1088/1742-6596/2071/1/012033.
15. *Santra Bikash & Ghosh Udit & Mukherjee Dipti.* Graph-based Modelling of Superpixels for Automatic Identification of Empty Shelves in Supermarkets // *Pattern Recognition.* – 2022. – 127 (2): 108627. – 10.1016/j.patcog.2022.108627.

16. Liu Song & Li Wanqing & Davis Stephen & Ritz Christian & Tian Hongda. Planogram Compliance Checking Based on Detection of Recurring Patterns // IEEE MultiMedia. – 2016. – Vol. 23. – P. 54-63. – 10.1109/MMUL.2016.19.
17. Баркалов С.А., Карнович М.А., Мусеев С.И. Метод анализа иерархий: подход, основанный на теории латентных переменных // Вестник ЮУрГУ. Серия: Компьютерные технологии, управление, радиоэлектроника. – 2022. – № 2.
18. Matai Rajesh & Singh Surya & Mittal M.L. Traveling Salesman Problem: an Overview of Applications, Formulations, and Solution Approaches. – 2010. – 10.5772/12909.
19. Cal Murat & Ekici Ali. Solving a Modified TSP Problem by a Greedy Heuristic for Cost Minimization // International Journal of Modeling and Optimization. – 2018. – 8. – P. 138-144. – 10.7763/IJMO.2018.V8.638.
20. Недвигин А.Р., Синецкий Р.М. Применение алгоритмов обработки изображений сверточных нейронных сетей в качестве основы для разработки системы поддержки принятия решения // Известия вузов. Северо-Кавказский регион. Серия: Технические науки. – 2025. – № 1.
21. Свидетельство о государственной регистрации программы для ЭВМ № 2025664496 Российская Федерация. Комплексная оценка доступности товаров с использованием технологий компьютерного зрения Ритейл-Диагност. Серверная часть. Версия 1.0 / А.Р. Недвигин, Р.М. Синецкий; Заявл. 15.05.2025; Запат. 04.06.2025.

REFERENCES

1. Tiwari Tanya & Tiwari Tanuj & Tiwari Sanjay. How Artificial Intelligence, Machine Learning and Deep Learning are Radically Different?, *International Journal of Advanced Research in Computer Science and Software Engineering*, 2018, 8(2):1, 10.23956/ijarsse.v8i2.569.
2. Mehwish Saqlain, Saddaf Rubab, Malik M. Khan, Nouman Ali, Shahzeb Ali. Hybrid Approach for Shelf Monitoring and Planogram Compliance (Hyb-SMPC) in Retails Using Deep Learning and Computer Vision, *Mathematical Problems in Engineering*, 2022, No. 1. Available at: <https://doi.org/10.1155/2022/4916818>.
3. Jorge Muñoz, Alonso Sanchez, Guillermo Kemper. End-to-end solution for automatic beverage stock detection in supermarkets based on image processing and convolutional neural networks, *International Journal of Cognitive Computing in Engineering*, 2024, No. 5, pp. 453-474. Available at: <https://doi.org/10.1016/j.ijcce.2024.09.001>.
4. Wei Yuchen, Tran Son, Xu Shuxiang, Kang Byeong, Springer Matthew. Deep Learning for Retail Product Recognition: Challenges and Techniques, *Computational Intelligence and Neuroscience*, 2020, 8875910, pp. 1-23. Available at: <https://doi.org/10.1155/2020/8875910>.
5. Gothai E. et al. Design features of grocery product recognition using deep learning, *Intelligent Automation and Soft Computing*, 2022, Vol. 34, No. 2, pp. 1231-1246.
6. Reynkhard Klette. Комп'ютерне зрєння. Теорія і алгоритми [Computer vision. Theory and algorithms]: transl. from Engl. by A.A. Slinkin. Moscow: DMK Press, 2019, 506 p.
7. Yolo Data Augmentation, 2025. Available at: <https://docs.ultralytics.com/guides/yolo-data-augmentation> (accessed 02 May 2025).
8. Surono Sugiyarto & Rivaldi Muhammad & Arrova Dewi Deshinta & Irsalinda Nursyiva. New Approach to Image Segmentation: U-Net Convolutional Network for Multiresolution CT Image Lung Segmentation, *Emerging Science Journal*, 2023, 7, pp. 498-506. 10.28991/ESJ-2023-07-02-014.
9. YOLOv9 Architecture Explained. Available at: <https://stunningvisionai.com/article/yolov9-architecture> (accessed 02 May 2025).
10. Desai Miral & Mewada Hiren & Pires Ivan & Roy Sparsh. Evaluating the Performance of the YOLO Object Detection Framework on COCO Dataset and Real-World Scenarios, *Procedia Computer Science*, 2024, 251, pp. 157-163. 10.1016/j.procs.2024.11.096.
11. Wang Decheng & Chen Xiangning & Yi Hui & Zhao Feng. Improvement of Non-Maximum Suppression in RGB-D Object Detection, *IEEE Access*, 2019, Vol. 7. 10.1109/ACCESS.2019.2945834.
12. Bondarenko V.A., Gessen P.A., Pavlova V.A., Sozinova M.V., Tupikov V.A. Algoritм obnaruzheniya ob"ektov dlya optiko-elektronnykh sistem s obucheniem v real'nom vremeni [Object detection algorithm for optoelectronic systems with online learning], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 1 (218).
13. Brink Kh., Richards D., Feverolf M. Mashinnoe obuchenie [Machine learning]. St. Petersburg: Piter, 2017, 336 p.
14. Said K & Jambek Asral. Analysis of Image Processing Using Morphological Erosion and Dilation, *Journal of Physics: Conference Series*, 2021, 2071(1):012033. 10.1088/1742-6596/2071/1/012033.

15. Santra Bikash & Ghosh Udit & Mukherjee Dipti. Graph-based Modelling of Superpixels for Automatic Identification of Empty Shelves in Supermarkets, *Pattern Recognition*, 2022, 127(2): 108627. 10.1016/j.patcog.2022.108627.
16. Liu Song & Li Wanqing & Davis Stephen & Ritz Christian & Tian Hongda. Planogram Compliance Checking Based on Detection of Recurring Patterns, *IEEE MultiMedia*, 2016, Vol. 23, pp. 54-63. 10.1109/MMUL.2016.19.
17. Barkalov S.A., Karpovich M.A., Moiseev S.I. Metod analiza ierarkhiy: podkhod, osnovanny na teorii latentnykh peremennykh [Hierarchy analysis method: an approach based on latent variable theory], *Vestnik YuUrGU. Seriya: Komp'yuternye tekhnologii, upravlenie, radioelektronika* [Bulletin of SUSU. Series: Computer Technologies, Control, Radio Electronics], 2022, No. 2.
18. Matai Rajesh & Singh Surya & Mittal M.L. Traveling Salesman Problem: an Overview of Applications, Formulations, and Solution Approaches, 2010. 10.5772/12909.
19. Cal Murat & Ekici Ali. Solving a Modified TSP Problem by a Greedy Heuristic for Cost Minimization, *International Journal of Modeling and Optimization*, 2018, 8, pp. 138-144. 10.7763/IJMO.2018.V8.638.
20. Nedvigina A.R., Sinetskiy R.M. Primenenie algoritmov obrabotki izobrazheniy svertochnykh neyronnykh setey v kachestve osnovy dlya razrabotki sistemy podderzhki prinyatiya resheniya [Application of image processing algorithms of convolutional neural networks as a basis for developing a decision support system], *Izvestiya vuzov. Severo-Kavkazskiy region. Seriya: Tekhnicheskie nauki* [Bulletin of Higher Educational Institutions. North Caucasus Region. Technical Sciences], 2025, No. 1.
21. Nedvigina A.R., Sinetskiy R.M. Svidetel'stvo o gosudarstvennoy registratsii programmy dlya EVM № 2025664496 Rossiyskaya Federatsiya. Kompleksnaya otsenka dostupnosti tovarov s ispol'zovaniem tekhnologii komp'yuternogo zreniya Riteyl-Diagnost. Servernaya chast'. Versiya 1.0 [Certificate of State Registration of a Computer Program No. 2025664496, Russian Federation. Comprehensive Assessment of Product Availability Using Computer Vision Technologies – Retail-Diagnost. Server Side. Version 1.0]; Filed on: 15.05.2025; Registered on: 04.06.2025.

Недвигин Артем Романович – Южно-Российский государственный политехнический университет (НПИ) имени М.И. Платова, e-mail: artemnedvigin@mail.ru; г. Новочеркасск, Россия; тел: 89185573541; кафедра «Программное обеспечение вычислительной техники»; аспирант.

Синецкий Роман Михайлович – Южно-Российский государственный политехнический университет (НПИ) имени М.И. Платова, e-mail: rmsin@srsru.ru; г. Новочеркасск, Россия; тел: 89289601848; кафедра «Программное обеспечение вычислительной техники»; к.т.н.; доцент.

Nedvigin Artem Romanovich – Platov South-Russian State Polytechnic University (NPI); e-mail: artemnedvigin@mail.ru; Novocherkassk, Russia; phone: +79185573541; the Department of Computer Software; graduate student.

Sinetskiy Roman Mikhailovich – Platov South-Russian State Polytechnic University (NPI); e-mail: rmsin@srsru.ru; Novocherkassk, Russia; phone: +79289601848; the Department of Computer Software; cand. of eng. sc.; associate professor.

УДК 681.327

DOI 10.18522/2311-3103-2026-1-33-42

П.О. Никашина

АЛГОРИТМ ОПРЕДЕЛЕНИЯ НЕЧЕТКОГО МНОЖЕСТВА СИЛЬНОЙ СВЯЗНОСТИ ПЕРИОДИЧЕСКОГО НЕЧЕТКОГО ГРАФА

Рассматривается метод определения сильной связности периодического нечеткого графа (PFG), который может быть применен для анализа динамических систем с учетом неопределенности и изменений во времени. Вводится понятие нечеткого множества сильной связности, которое позволяет оценить степень достижимости между вершинами графа за определенное количество тактов времени. Основное внимание уделяется разработке алгоритма для нахождения данного множества, что является важным инструментом для анализа связности в сложных системах. Статья начинается с обзора существующих подходов к анализу связности в нечетких графах, подчеркивая необходимость учета временных и нечетких параметров. Основная часть работы посвящена описанию ключевых понятий и определений, связанных с периодическими не-

четкими графами. Вводятся понятия нечеткого пути, времени и степени достижимости, а также нечеткого множества достижимости. Предлагается алгоритм для нахождения нечеткого множества достижимости, основанный на волновом методе, который позволяет определить степень и время достижимости между вершинами графа. Далее вводится понятие нечеткого множества сильной связности PFG и предлагается алгоритм для его определения. В качестве примера рассматривается конкретный PFG, для которого вычисляется нечеткое множество сильной связности. Предложенный метод может быть полезен для анализа и оптимизации процессов в динамических системах, где модель взаимодействий представлена периодическим нечетким графом. В будущем планируется исследовать вопросы, связанные с нахождением дискретного времени достижимости между вершинами при заданной степени достижимости, что расширит возможности применения данного подхода в различных областях.

Периодический нечеткий граф; множество достижимости; нечеткое множество сильной связности.

P.O. Nikashina

ALGORITHM FOR DETERMINING A STRONGLY CONNECTED FUZZY SET OF A PERIODIC FUZZY GRAPH

The article discusses a method for determining the strong connectivity of a periodic fuzzy graph (PFG), which can be used to make decisions in emergency situations such as evacuation. The concept of a fuzzy set of strong connectivity is introduced. The main attention is paid to the development of a method for finding a fuzzy set of strong connectivity, which makes it possible to determine the degree of reachability between the vertices of a graph in a certain number of time cycles. The paper begins with a review of existing approaches to connectivity analysis in fuzzy graphs, emphasizing the need to consider time and fuzzy parameters. The main part of the paper is devoted to the description of key concepts and definitions related to periodic fuzzy graphs. The concepts of fuzzy path, time and degree of reachability, as well as fuzzy set of reachability are introduced. An algorithm for finding a fuzzy set of reachability based on the wave method is proposed, which allows determining the degree and time of reachability between graph vertices. Next, the concept of fuzzy set of strong connectivity PFG is introduced and an algorithm for its determination is proposed. As an example, a specific PFG is considered for which the fuzzy set of strong connectivity is calculated. The proposed method can be useful for choosing a movement strategy during evacuation, especially in conditions where the territory model is represented by a periodic fuzzy graph. In the future, it is planned to explore issues related to finding a discrete-time reachability between vertices with a given degree of reachability.

Periodic fuzzy graph; reachability set; strongly connected fuzzy set.

Введение. В современном мире, где динамические системы и сложные сети играют ключевую роль в различных областях, таких как транспортные системы, логистика, управление потоками данных и даже социальные сети, возникает необходимость в разработке методов, способных учитывать неопределенность и временные изменения. Одним из таких инструментов являются периодические нечеткие графы (PFG), которые позволяют моделировать системы с изменяющимися во времени параметрами и нечеткими связями между элементами [1, 2].

Периодические нечеткие графы представляют собой мощный математический аппарат для анализа и оптимизации процессов, где важно учитывать не только структуру связей, но и их динамику. В таких графах вершины могут представлять различные объекты (например, узлы транспортной сети, точки сбора данных или элементы инфраструктуры), а ребра – связи между ними, которые могут изменяться во времени и иметь различную степень "силы" или "надежности". Это делает PFG особенно полезными для задач, где требуется учитывать не только текущее состояние системы, но и ее эволюцию во времени.

В данной статье рассматривается метод определения сильной связности периодического нечеткого графа, который может быть применен для анализа и оптимизации различных процессов, связанных с перемещением или передачей данных в динамических системах. Основное внимание уделяется разработке алгоритма, позволяющего определить степень достижимости между вершинами графа за определенное количество тактов

времени. Этот метод может быть полезен для решения задач, где требуется оценить, насколько эффективно можно перемещаться между различными точками системы в условиях неопределенности и изменяющихся условий.

Предложенный подход может быть применен в различных областях, таких как:

- ♦ Транспортные системы. Оптимизация маршрутов в условиях изменяющейся загрузки дорог или других факторов, влияющих на пропускную способность. В работе [3] представлены различные методы и инструменты для поддержки эффективного планирования, однако отсутствует общая модель, которая могла бы учитывать все необходимые параметры.

- ♦ Логистика. Планирование перемещения грузов в условиях неопределенности, например, при изменении доступности транспортных узлов или маршрутов. В [4] рассмотрены этапы планирования, но отсутствует метод оценки общей информации о ситуации, что затрудняет принятие решений.

- ♦ Управление потоками данных. Анализ и оптимизация передачи информации в сетях, где связи между узлами могут изменяться во времени. Исследования, проведенные в [5], определили ключевые задачи для разработки планов, включая определение прогнозируемых параметров и сценариев, а также оптимизацию стратегий.

- ♦ Социальные сети. Изучение динамики взаимодействия между пользователями, где связи могут быть нечеткими и изменяться в зависимости от времени. В [6] представлена программа моделирования, которая может быть адаптирована для анализа таких систем.

В статье вводится понятие нечеткого множества сильной связности, которое позволяет определить, насколько устойчивы связи между элементами системы в условиях неопределенности и временных изменений. Предложенный алгоритм, основанный на волновом методе, позволяет эффективно вычислять степень достижимости между вершинами графа, что делает его полезным инструментом для анализа и оптимизации сложных динамических систем.

Таким образом, данный метод может быть использован не только для решения задач, связанных с чрезвычайными ситуациями, как например, в некоторых работах можно увидеть предложенные методы и инструменты для моделирования эвакуации, которые можно разделить на два типа: модели, ориентированные на конкретные бедствия [7, 8], и модели поддержки эвакуации [9–11]. Но и для широкого спектра приложений, где требуется учитывать неопределенность и динамику изменений в системах. В будущем планируется исследовать вопросы, связанные с нахождением дискретного времени достижимости между вершинами при заданной степени достижимости, что позволит еще более точно моделировать и оптимизировать процессы в динамических системах.

1. Основные понятия и определения периодического нечеткого графа. Приведем некоторые понятия и определения периодического нечеткого графа (PFG): нечеткий путь, время и степень достижимости, нечеткое множество достижимости [12, 13].

Периодическим нечетким графом называется нечеткий темпоральный граф $\tilde{G} = (V, \tilde{U}_t, T)$, где V – множество вершин графа с числом вершин n ; $T = \{1, 2, \dots, N\}$ – множество натуральных чисел, определяющих дискретное время (horizon); $\tilde{U}_t = \{< \mu_t(x_i, x_j) / (x_i, x_j) >\}$ – нечеткое множество ребер, где $x_i, x_j \in V$, $\mu_t(x_i, x_j) \in [0, 1]$ – значение функции принадлежности μ_t для ребра (x_i, x_j) в момент времени $t \in T$, у которого дискретное время обладает свойством цикличности.

Иными словами, в PFG после момента времени N опять наступает момент «1».

Переход от одного момента времени (t) в следующий момент ($t+1$) назовем *одним тактом*.

Вершина x_j называется *нечетко смежной* вершине x_i если справедливо $(\exists t \in T) \mu_t(x_i, x_j) > 0$.

Нечетким путем $\tilde{L}(x_i, x_j)$ PFG $\tilde{G}$ называется последовательность нечетких дуг, ведущих из вершины x_i в вершину x_j , в которой конечная вершина всякой дуги (кроме последней), является начальной вершиной следующей дуги: $\tilde{L}(x_i, x_j) = < \mu_{t_1}(x_i, x_1) /$

$(x_i, x_1) >, \dots, < \mu_{t_j}(x_{j-1}, x_j) / (x_{j-1}, x_j) >$, для которой выполняются условия $\mu_{t_1}(x_i, x_1) > 0$, $\mu_{t_2}(x_2, x_3) > 0, \dots, \mu_{t_k}(x_{k-1}, x_k) > 0$. Причем, эта последовательность не зависит от моментов времени $t_1, t_2, \dots, t_k \in T$.

Вершину x_j является *нечетко достижимой* из вершины x_i в PFG $\tilde{G}$ если существует ориентированный нечеткий путь $\tilde{L}(x_i, x_j)$.

Время достижимости $\tau(x_i, x_j)$ есть число тактов, которое необходимо затратить, чтобы из вершины x_i достичь вершину x_j по пути $\tilde{L}(x_i, x_j)$.

Степень достижимости $\mu_L(x_i, x_j)$ из вершины x_i в вершину x_j по пути $\tilde{L}(x_i, x_j)$ есть наименьшее значение функции принадлежности дуги, входящей в этот путь:

$$\mu_L(x_i, x_j) = \mu_{t_1}(x_i, x_1) \& \mu_{t_2}(x_1, x_2) \& \dots \& \mu_{t_j}(x_{j-1}, x_j).$$

Пусть μ_τ – наибольшая степень достижимости вершины x_f из вершины x_s за число тактов τ .

Нечеткое множество достижимости есть множество:

$$\tilde{\tau}(x_s, x_f) = \{ < \mu_\tau, \tau >, \mu_\tau \in [0, 1], \tau \in 1, 2, \dots, N \times (n - 1) \}.$$

Для нахождения нечеткого множества достижимости предлагается использовать алгоритм, основанный на идее волнового алгоритма, предназначенного для нахождения кратчайших путей в четких графах:

1°. Вершине x_s присваиваем нечеткое множество времени достижимости в виде пары $\tilde{\tau}_s = \{ < 1, (0, 0) > \}$. Здесь 1 – степень достижимости самой вершины x_s в циклическое время $(0, 0)$. Всем остальным вершинам присваиваем $\tilde{\tau}_j = \{ < 0, (N \times (n - 1), N) > \}$.

2°. Каждой вершине x_i смежной с вершиной x_s по степеням $\mu_1(x_s, x_i), \mu_2(x_s, x_i), \dots, \mu_N(x_s, x_i)$ присваиваем нечеткое множество времени достижимости $\{ < \mu_{i1}, (0, 1) >, < \mu_{i2}, (0, 2) >, \dots, < \mu_{iN}, (0, N) > \}$. Здесь $(\forall j = \overline{1, N}) [\mu_{ij} = \mu_j(x_s, x_i)]$. Полученное множество дописываем в $\tilde{\tau}_i$.

3°. Вершине x_{i+1} смежной с вершиной x_i по степеням $\mu_1(x_i, x_{i+1}), \mu_2(x_i, x_{i+1}), \dots, \mu_N(x_i, x_{i+1})$ вычисляем нечеткое множество времени достижимости

$$\begin{aligned} & \{ < \mu_{i1} \& \mu_2(x_i, x_{i+1}), (0, 2) >, < \mu_{i2} \& \mu_3(x_i, x_{i+1}), (0, 3) >, \dots, < \\ & \mu_{iN-1} \& \mu_N(x_i, x_{i+1}), (0, N) >, < \mu_{i1} \& \mu_1(x_i, x_{i+1}), (1, 1) >, \dots, < \\ & \mu_{iN} \& \mu_N(x_i, x_{i+1}), (1, N) > \}. \end{aligned}$$

Полученное множество также дописываем в $\tilde{\tau}_{i+1}$.

Внутри вычисленных множеств $\tilde{\tau}_i$ уберем те пары $< \text{степень}, \text{время} >$, у которых время «не меньше», а степень «не больше» чем у некоторой оставшейся во множестве пары. Для этого воспользуемся правилом поглощения:

$$\mu_1 \geq \mu_2 \& t_1 \leq t_2 \rightarrow < \mu_1, t_1 > \& < \mu_2, t_2 > = < \mu_1, t_1 >. \quad (1)$$

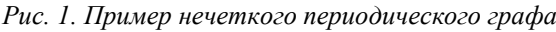
Например:

$$< 0.4, (0, 4) > \& < 0.6, (0, 2) > = < 0.6, (0, 2) >;$$

$$< 0.8, (1, 3) > \& < 0.8, (3, 3) > = < 0.8, (1, 3) >.$$

4°. Вычисленное нечеткое множество $\tilde{\tau}_f$ определит время и степень достижимости вершины x_f из вершины x_s . Согласно выражению (1) определяем $\tilde{\tau}(x_s, x_f)$.

Пример 4. Рассмотрим пример нахождения достижимости вершины x_5 из вершины x_1 в PFG $\tilde{G}$, приведенном на рис. 1.



- ♦ вершине x_1 прис

- $\tilde{t}_i = \{< 0, (12,3) >, i = \overline{2,5}$. Здесь $N \times (n - 1) = 3 \times (5 - 1) = 12$;
- ♦ вершинам x_2 и x_4 после поглощения пары $< 0, (12,3) >$ приписываются множества $\tilde{t}_2 = \{< 0.8, (0,1) >, < 1, (0,2) >\}$ и $\tilde{t}_4 = \{< 0.8, (0,1) >, < 1, (0,2) >\}$;
 - ♦ вершине x_3 присваивается $\tilde{t}_3 = \{< 0.7, (0,2) >, < 0.8, (0,3) >, < 0.9, (0,3) >\}$.
- После поглощения, согласно правилу (2), остается $\tilde{t}_3 = \{< 0.7, (0,2) >, < 0.9, (0,3) >\}$.
- ♦ вершине x_5 присваивается $\tilde{t}_5 = \{< 0.6, (0,3) >, < 0.7, (1,2) >, < 0.9, (1,3) >, < 0.8, (0,2) >, < 0.9, (1,2) >\}$. После поглощения, согласно правилу (2), остается $\tilde{t}_5 = \{< 0.8, (0,2) >, < 0.9, (1,2) >\}$.

Таким образом были исследованы основные принципы периодических нечётких графов: их формальное описание, построение и свойства. Разобраны структурные элементы – вершины, рёбра с нечёткими весами и временные циклы, необходимые для работы с динамическими системами.

Предложенный алгоритм нахождения множества достижимости основан на волновом методе, который позволяет определить минимальное время и степень достижимости для каждой вершины. Этот метод может быть использован в различных задачах, связанных с эвакуацией, транспортными потоками, анализом сетевых структур и принятием решений в условиях неопределенности.

2. Нечеткое множество сильной связанности PFG

Определение 1. PFG $\tilde{G}$ назовем *нечетко сильно связным*, если между любыми двумя вершинами графа существует путь со степенью достижимости $\neq 0$ за некоторое число тактов $\tau \in [1, \dots, N \times (n - 1)]$. В противном случае PFG $\tilde{G}$ не обладает сильной связностью.

Определение 1. PFG $\tilde{G}$ назовем нечетко сильно связным, если

Произвольный PFG $\tilde{G}$ может быть не сильно связным, а также быть сильно связным с разной степенью за различное число τ . В связи в этом для анализа PFG $\tilde{G}$ возникают следующие задачи:

1. Определить, является ли PFG $\tilde{G}$ сильно связным.
2. Если да, то:
 - 2.1. Определить степень сильной связности и соответствующее время достижимости;

2.2. Определить наибольшую степень сильной связности за наименьшее возможное для этого время достижимости.

Пусть $\mu_t(x_i, x_j)$ – наибольшая степень достижимости вершины x_j из вершины x_i за число тактов $t \leq \tau$.

Обозначим через $\mu_G(t) = \min_{\forall x_i, x_j \in V} \{\mu_t(x_i, x_j)\}$ степень сильной связности PFG $\tilde{G}$ за время t .

Определение 2. Нечетким множеством сильной связности PFG $\tilde{G}$ назовем множество

$$S(\tilde{G}) = \{ \langle \mu_G(t)/t \rangle \mid t \in [1, N \times (n-1)] \}.$$

Справедливо следующее свойство:

- $t_1 < t_2 \rightarrow \mu_G(t_1) \leq \mu_G(t_2)$.

Нечеткое множество сильной связности позволяет решить указанные выше задачи:

- ♦ Если $S(\tilde{G}) = \emptyset$, то PFG $\tilde{G}$ не является нечетко сильно связным;
- ♦ Если $S(\tilde{G}) \neq \emptyset$, то:

а) наименьшее значение t_{min} в $S(\tilde{G})$ определит наименьшее время, при котором PFG $\tilde{G}$ обладает сильной связностью со степенью $\mu_G(t_{min})$, то есть для числа тактов $\tau < t_{min}$ в графе существует хотя бы одна пара вершин, между которыми нет нечеткого пути.

б) величина $\mu^0 = \max_t \{\mu_G(t)\}$ определит наибольшую степень сильной связности, а величина $t^0 = \min_t \{t \mid \mu_G(t) = \mu^0\}$ определит наименьшее время достижимости, при котором PFG $\tilde{G}$ обладает этой наибольшей степенью сильной связности μ^0 .

Для определения нечеткого множества сильной связности предлагается следующий алгоритм:

1⁰. $t := 1$.

2⁰. Определяем значение $\mu_t = \min_{\forall x_i, x_j \in V} \{r_{ij}\}$, $i, j \in \overline{1, n}$. Здесь r_{ij} – элементы матрицы нечеткой достижимости $R = ||r_{ij}||$.

3⁰. Если $\mu_t \neq 0$, или $t \neq 1$ & $\mu_t \neq \mu_{t-1}$ то пара $\langle \mu_t/t \rangle$ включается в $S(\tilde{G})$.

4⁰. $t := t + 1$.

5⁰. Если $t \leq N \times (n-1)$ или $\mu_t^0 = 1$ то 2⁰, иначе 6⁰.

6⁰. Конец.

Таким образом было введено и формализовано понятие нечеткого множества сильной связности для периодического нечеткого графа (PFG). Рассмотрены ключевые характеристики сильной связности в условиях нечеткости и периодичности временных изменений, что позволяет анализировать устойчивость структуры графа и его применимость в задачах эвакуации, транспортных систем и сетевого анализа.

Было показано, что нечеткая сильная связность определяется не только наличием путей между вершинами, но и степенью достижимости за определенное число тактов. Введено определение степени сильной связности и разработан алгоритм вычисления нечеткого множества сильной связности, который позволяет:

- ♦ определить, является ли PFG сильно связным;
- ♦ найти минимальное время и соответствующую степень связности;
- ♦ выявить наибольшую степень сильной связности за наименьшее возможное время.

Рассмотренный алгоритм, основанный на анализе матрицы нечеткой достижимости, позволяет эффективно решать задачи оценки связности графа и поиска оптимальных стратегий перемещения в динамических системах.

Результаты примера демонстрируют, что степень сильной связности PFG может изменяться в зависимости от числа тактов, а ее максимальное значение достигается не всегда за минимальное время. Это подтверждает, что учет временного аспекта играет важную роль при анализе структуры графа.

3. Практическое применение алгоритма. Рассмотрим первый пример определения нечеткого множества сильной связности для PFG, представленного на рис. 1.

Элементы матрицы $R = ||r_{ij}||$ степеней достижимости $\tilde{r}(x_i, x_j)$, $i, j = \overline{1, 5}$ между вершинами графа приведены в табл. 1:

Таблица 1

Степени достижимости $\tilde{t}$

	x_1	x_2	x_3	x_4	x_5
x_1	<1,0>	<0.8,1>, <1,2>	<0.7,2>, <0.9,3>	<0.8,1>, <1,3>	<0.8,2>, <0.9,5>
x_2	<0.8,7>, <0.9,8>	<1,0>	<0.7,2>, <0.9,3>	<0.9,5>	<0.6,3>, <0.9,4>
x_3	<0.8,4>, <1,5>	<0.8,5>, <1,6>	<1,0>	<1,2>	<1,1>
x_4	<1,2>	<0.8,2>, <1,5>	<0.7,2>, <0.8,3>, <0.9,6>	<1,0>	<0.9,2>
x_5	<0.6,2>, <0.8,4>, <1,5>	<0.6,4>, <0.8,7>, <1,8>	<0.6,3>, <0.7,5>, <0.9,9>	<0.6,1>, <1,2>	<1,0>

Согласно алгоритма для $t=1,2,\dots,6$ определяем величины: $\mu_t = 0$; $\mu_7 = 0.7$, $\mu_8 = 0.7$, $\mu_9 = \mu_{10} = \dots = 0.9$. Откуда получаем, что:

$$S(\tilde{G}) = \{< 0.7/7 >, < 0.9/9 >\}.$$

Таким образом, PFG $\tilde{G}$ имеет степень сильной связности 0.7 за наименьшее число тактов 7 и наибольшую степень связности 0.9 за число тактов 9.

Заключение. В данной статье представлена разработка метода определения сильной связности для периодического нечеткого графа (PFG), который может быть применен для анализа динамических систем с учетом неопределенности и изменений во времени. Рассмотрены основные понятия, связанные с периодическими нечеткими графами, такие как нечеткий путь, степень достижимости и методы их вычисления, а также предложен алгоритм нахождения нечеткого множества сильной связности, который является основой предложенной методики.

В работе изложен теоретический аппарат для анализа периодических нечетких графов, включая формализацию ключевых определений, таких как нечеткий путь, время достижимости и нечеткое множество достижимости. Эти понятия играют важную роль в создании математических моделей, способных учитывать изменения в графах с течением времени и неопределенность данных. Одним из ключевых аспектов исследования стало введение понятия нечеткого множества сильной связности для периодического нечеткого графа, что расширяет возможности анализа связности в динамических и нечетких условиях.

Предложенный алгоритм для нахождения нечеткого множества сильной связности основан на методах волнового поиска, адаптированных для работы с нечеткими величинами и цикличностью во времени. Алгоритм детально описан и проиллюстрирован на примере, что позволяет наглядно продемонстрировать процесс вычисления степени сильной связности и ее зависимость от времени.

Метод нахождения нечеткого множества сильной связности обладает широким потенциалом для применения в различных областях. Он может быть использован для анализа и оптимизации сетевых структур, моделирования транспортных потоков в условиях неопределенности, анализа устойчивости инфраструктур, а также для решения задач, связанных с динамическими системами, где требуется учитывать временные изменения и нечеткость данных. Например, предложенный подход может быть полезен в логистике для планирования маршрутов, в телекоммуникациях для анализа устойчивости сетей, в энергетике для управления распределенными системами, а также в социальных и экономических исследованиях для моделирования взаимодействий в сложных системах.

В дальнейшем планируется исследовать вопросы, связанные с оптимизацией алгоритма и его применением для решения задач, требующих учета дискретного времени достижимости при заданной степени связности. Это позволит расширить область применения метода и повысить его эффективность в практических задачах.

Подтверждения. Работа выполнена при поддержке Министерства транспорта Российской Федерации в рамках проекта № 103-00001-26-00 «Разработка и исследование методологии интеллектуального геоинформационного моделирования транспортных процессов в условиях неполноты информации».

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Никашина П.О. Определение максимального потока в нечетком периодическом графе // Известия ЮФУ. Технические науки. – 2024. – № 4 (240). – С. 31-40.
2. Никашина П.О., Боженьюк А.В. Оптимизация транспортного потока на основе периодических нечетких графов // Известия ЮФУ. Технические науки. – 2023. – № 4 (234). – С. 97-108.
3. Shaw D., Albores P., Anson S., Kailiponi P., Nagarajan M., Tissington P., and Hart T. Evacuation Responsiveness by Government Organisations (ERGO) // Evacuation Preparedness Assessment Workbook. Technical report. Aston CRISIS Center, 2011.
4. Lumbroso D., and Vinet F. Tools to Improve the Production of Emergency Plans for Floods: Are They Being Used by the People that Need Them? // Journal of Contingencies and Crisis Management. – 2012. – Vol. 20. – P. 149-165.
5. Hissel M., François H., and Xiao J.J. Support for Preventive Mass Evacuation Planning in Urban Areas // IET Conference Publications. – 2011. – 582. – P. 159-165.
6. Chiu Y., and Liu H.X. Emergency Evacuation, Dynamique Transportation Model, Spring Street, NY 10013, USA: Springer Science Buisness Media, LLC, 2008.
7. Bayram V. Optimization models for large scale network evacuation planning and management: A literature review // Surveys in Operations Research and Management Science. – 2016. – Vol. 21 (2). – P. 63-84.
8. Gao Z., Qu Y., Li X., Long J., and Huang H.-J. Simulating the dynamic escape process in large public places // Operations Research. – 2014. – Vol. 62 (6). – P. 1344-1357.
9. Lazo J.K., Waldman D.M., Morrow B.H., and Thacher J.A. Household evacuation decision making and the benefits of improved hurricane forecasting: Developing a framework for assessment // Weather and Forecasting. – 2010. – Vol. 25 (1). – P. 207-219.
10. Simonovic S.P., and Ahmad S. Computer-based model for flood evacuation emergency planning // Natural hazards. – 2005. – Vol. 34 (1). – P. 25-51.
11. Dash N., and Gladwin H. Evacuation decision making and behavioral responses: Individual and household // Natural Hazards Review. – 2007. – Vol. 8 (3). – P. 69-77.
12. Bozhenyuk A., Gorbachev S., and Knyazeva M. Finding Fuzzy Sets of Bases and Antibases of Periodic Fuzzy Graph // Lecture Notes in Networks and Systems (LNNS). – 2024. – Vol. 1089 (2). – P. 767-774.
13. Bozhenyuk A., Knyazeva M., Kosenko O., and Rozenberg I. Strong Connectivity Definition of Periodic Fuzzy Graph // Lecture Notes in Networks and Systems (LNNS). – 2023. – Vol. 758. – P. 168-174.
14. Lessan J., and Kim A.M. Planning evacuation orders under evacuee compliance uncertainty // Safety Science. – 2022. – 156, 105894.
15. Stepanov A., and Smith M.J. Multi-objective evacuation routing in transportation networks // European Journal of Operational Research. – 2009. – Vol. 198 (2). – P. 435-446.
16. Lindell M., and Prater C. Critical behavioral assumptions in evacuation time estimate analysis for private vehicles: Examples from hurricane research and planning // Journal of Urban Planning and Development. – 2007. – Vol. 133 (1). – P. 18-29.
17. Bretschneider S. Mathematical models for evacuation planning in urban areas // Lecture Notes in Economics and Mathematical Systems. – Vol. 659. Verlag Berlin Heidelberg, Springer, 2012.
18. Hissel F. Methodology for the implementation of mass evacuation plans, CEMEF, France, Compiègne, 2011.
19. Kailiponi P. Analyzing evacuation decision using Multi-Attribute Utility Theory (MAUT) // Procedia Engineering. – 2010. – Vol. 3. – P. 163-174.
20. Regnier E. Public evacuation decision and hurricane track uncertainty // Management Science. – 2008. – Vol. 54 (1). – P. 16-28.
21. Agumya A., and Hunter G.J. Responding to the consequences of uncertainty in geographical data // International Journal of Geographical Information Science. – 2002. – Vol. 16 (5). – P. 405-417.
22. Kunz M., Gret-Regamey A., and Hurni L. Visualization of uncertainty in natural hazards assessments using an interactive cartographic information system // Natural Hazards. – 2011. – Vol. 59 (3). – P. 1735-1751.
23. Kacprzyk J., Zadrozny S., Nurmi H., and Bozhenyuk A. Towards Innovation Focused Fuzzy Decision Making by Consensus // In: Proceedings of IEEE Int. Conf. on Fuzzy Systems. – 2021. – P. 256-268.

24. Bozhenyuk A., Gorbachev S., and Knyazeva M. Finding Fuzzy Sets of Bases and Antibases of Periodic Fuzzy Graph // *Lecture Notes in Networks and Systems (LNNS)*. – 2024. – Vol. 1089 (2). – P. 767-774.
25. Bozhenyuk A., Knyazeva M., Kosenko O., and Rozenberg I. Strong Connectivity Definition of Periodic Fuzzy Graph // *Lecture Notes in Networks and Systems (LNNS)*. – 2023. – Vol. 758. – P. 168-174.

REFERENCES

1. Nikashina P.O. Opredelenie maksimal'nogo potoka v nechetkom periodicheskom grafe [Determining the maximum flow in a fuzzy periodic graph], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2024, No. 4 (240), pp. 31-40.
2. Nikashina P.O., Bozhenyuk A.V. Optimizatsiya transportnogo potoka na osnove periodicheskikh nechetkikh grafov [Optimization of traffic flow based on periodic fuzzy graphs], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2023, No. 4 (234), pp. 97-108.
3. Shaw D., Albores P., Anson S., Kailiponi P., Nagarajan M., Tissington P., and Hart T. Evacuation Responsiveness by Government Organisations (ERGO), Evacuation Preparedness Assessment Workbook. Technical report. Aston CRISIS Center, 2011.
4. Lumbroso D., and Vinet F. Tools to Improve the Production of Emergency Plans for Floods: Are They Being Used by the People that Need Them?, *Journal of Contingencies and Crisis Management*, 2012, Vol. 20, pp. 149-165.
5. Hissel M., François H., and Xiao J.J. Support for Preventive Mass Evacuation Planning in Urban Areas, *IET Conference Publications*, 2011, 582, pp. 159-165.
6. Chiu Y., and Liu H.X. Emergency Evacuation, Dynamique Transportation Model, Spring Street, NY 10013, USA: Springer Science Buisiness Media, LLC, 2008.
7. Bayram V. Optimization models for large scale network evacuation planning and management: A literature review, *Surveys in Operations Research and Management Science*, 2016, Vol. 21 (2), pp. 63-84.
8. Gao Z., Qu Y., Li X., Long J., and Huang H.-J. Simulating the dynamic escape process in large public places, *Operations Research*, 2014, Vol. 62 (6), pp. 1344-1357.
9. Lazo J.K., Waldman D.M., Morrow B.H., and Thacher J.A. Household evacuation decision making and the benefits of improved hurricane forecasting: Developing a framework for assessment, *Weather and Forecasting*, 2010, Vol. 25 (1), pp. 207-219.
10. Simonovic S.P., and Ahmad S. Computer-based model for flood evacuation emergency planning, *Natural hazards*, 2005, Vol. 34 (1), pp. 25-51.
11. Dash N., and Gladwin H. Evacuation decision making and behavioral responses: Individual and household, *Natural Hazards Review*, 2007, Vol. 8 (3), pp. 69-77.
12. Bozhenyuk A., Gorbachev S., and Knyazeva M. Finding Fuzzy Sets of Bases and Antibases of Periodic Fuzzy Graph, *Lecture Notes in Networks and Systems (LNNS)*, 2024, Vol. 1089 (2), pp. 767-774.
13. Bozhenyuk A., Knyazeva M., Kosenko O., and Rozenberg I. Strong Connectivity Definition of Periodic Fuzzy Graph, *Lecture Notes in Networks and Systems (LNNS)*, 2023, Vol. 758, pp. 168-174.
14. Lessan J., and Kim A.M. Planning evacuation orders under evacuee compliance uncertainty, *Safety Science*, 2022, 156, 105894.
15. Stepanov A., and Smith M.J. Multi-objective evacuation routing in transportation networks, *European Journal of Operational Research*, 2009, Vol. 198 (2), pp. 435-446.
16. Lindell M., and Prater C. Critical behavioral assumptions in evacuation time estimate analysis for private vehicles: Examples from hurricane research and planning, *Journal of Urban Planning and Development*, 2007, Vol. 133 (1), pp. 18-29.
17. Bretschneider S. Mathematical models for evacuation planning in urban areas, *Lecture Notes in Economics and Mathematical Systems*, Vol. 659. Verlag Berlin Heidelberg, Springer, 2012.
18. Hissel F. Methodology for the implementation of mass evacuation plans, CEMEF, France, Compiègne, 2011.
19. Kailiponi P. Analyzing evacuation decision using Multi-Attribute Utility Theory (MAUT), *Procedia Engineering*, 2010, Vol. 3, pp. 163-174.
20. Regnier E. Public evacuation decision and hurricane track uncertainty, *Management Science*, 2008, Vol. 54 (1), pp. 16-28.
21. Agumya A., and Hunter G.J. Responding to the consequences of uncertainty in geographical data, *International Journal of Geographical Information Science*, 2002, Vol. 16 (5), pp. 405-417.
22. Kunz M., Gret-Regamey A., and Hurni L. Visualization of uncertainty in natural hazards assessments using an interactive cartographic information system, *Natural Hazards*, 2011, Vol. 59 (3), pp. 1735-1751.
23. Kacprzyk J., Zadrozny S., Nurmi H., and Bozhenyuk A. Towards Innovation Focused Fuzzy Decision Making by Consensus, In: *Proceedings of IEEE Int. Conf. on Fuzzy Systems*, 2021, pp. 256-268.

24. Bozhenyuk A., Gorbachev S., and Knyazeva M. Finding Fuzzy Sets of Bases and Antibases of Periodic Fuzzy Graph, *Lecture Notes in Networks and Systems (LNNS)*, 2024, Vol. 1089 (2), pp. 767-774.
25. Bozhenyuk A., Knyazeva M., Kosenko O., and Rozenberg I. Strong Connectivity Definition of Periodic Fuzzy Graph, *Lecture Notes in Networks and Systems (LNNS)*, 2023, Vol. 758, pp. 168-174.

Никашина Полина Олеговна – Южный федеральный университет; e-mail: nikashina@sfedu.ru; г. Таганрог, Россия; тел.: +79381453823; кафедра информационно-аналитических систем безопасности имени профессора Берштейна Леонида Самойловича; аспирант.

Nikashina Polina Olegovna – Southern Federal University; e-mail: nikashina@sfedu.ru; Taganrog, Russia; phone: +79381453823; the Department of Information and Analytical Security Systems named after Professor Bershtein Leonid Samoylovich; graduate student.

УДК 004.023

DOI 10.18522/2311-3103-2026-1-42-52

М.В. Пикалов

ВЛИЯНИЕ МЕТОДА ВЫБОРКИ НА ОЦЕНКУ ХАРАКТЕРИСТИК В АНАЛИЗЕ ЛАНДШАФТА ПОИСКА

Методы анализа ландшафта поиска (ELA) оценивают ландшафт приспособленности задачи в виде числовых дескрипторов, часто используемых для рекомендации оптимальных параметров алгоритма. Данное исследование изучает влияние методов и объёма выборки на аппроксимацию признаков ELA и последующую производительность моделей машинного обучения. Исследование демонстрирует, что эти аппроксимированные признаки не являются абсолютными характеристиками ландшафта, а существенно зависят от метода, использованного для генерации точек выборки. Хотя увеличение объёма выборки снижает дисперсию оценок признаков, сам выбор метода семплирования вносит значительное смещение, приводя к статистически различным значениям признаков для таких методов, как вихрь Мерсенна, латинский гиперкуб (LHS) и последовательности Фауре. Ключевой эксперимент включал предсказание параметров настраиваемой задачи W-model с использованием регрессионных моделей, обученных на признаках ELA. Результаты показали, что модели, обученные и протестированные на данных, полученных одним и тем же методом выборки, показали наилучшие результаты, что подчеркивает отсутствие совместимости между различными методами выборки. Примечательно, что квазислучайные последовательности Фауре давали наименьшую ошибку регрессии, превосходя распространенные методы, такие как равномерное случайное распределение и LHS. Более того, перекрёстная проверка выявила, что модели, особенно обученные на последовательностях Фауре, демонстрировали значительное падение производительности при тестировании на данных, полученных любым другим методом, что подтверждает: стратегия выборки накладывает специфический «отпечаток» на данные признаков. Полученные результаты ставят под сомнение стандартное использование распространенных методов выборки в ELA. Точность моделей машинного обучения для выбора и настройки алгоритмов сильно зависит от метода выборки, используемой для извлечения признаков. Следовательно, для обеспечения точности критически важно соблюдать согласованность между методами выборки, используемыми на этапах обучения модели и её практического применения. Высокая производительность последовательностей Фауре указывает на то, что подобные последовательности являются перспективным направлением для будущих исследований в области создания более надёжных и точных моделей на основе ELA.

Эволюционные алгоритмы; настройка параметров; анализ ландшафта поиска.

M.V. Pikalov

IMPACT OF SAMPLING TECHNIQUES ON EXPLORATORY LANDSCAPE ANALYSIS

Exploratory Landscape Analysis (ELA) features are numerical descriptors of a problem's fitness landscape, often used to recommend optimal algorithm parameters. This study investigates the critical impact of sampling methods and sample size on the approximation of ELA features and the subsequent performance of machine learning models. The research demonstrates that these feature approximations are not absolute characteristics of the landscape but are significantly influenced by the method used to

generate the sample points. While increasing the sample size reduces the variance of feature estimates, the choice of sampling strategy itself introduces substantial bias, leading to statistically different feature values across methods like Mersenne Twister, Latin Hypercube Sampling (LHS), and Faure sequences. The core experiment involved predicting the parameters of the tunable W-model problem using regression models trained on ELA features. The results showed that models trained and tested on data from the same sampling method performed best, highlighting a lack of interoperability between different sampling techniques. Notably, the Faure quasirandom sequences consistently yielded the lowest regression error, outperforming common methods like uniform random sampling and LHS. Furthermore, cross-sampling validation revealed that models, especially those trained on Faure sequences, suffered a significant performance drop when tested on data from any other method, confirming that the sampling strategy imparts a specific "fingerprint" on the feature data. In conclusion, the findings challenge the default use of common sampling methods in ELA. The accuracy of machine learning models for algorithm selection and configuration is highly sensitive to the sampling strategy employed for feature extraction. Therefore, ensuring consistency between the sampling methods used during model training and application is crucial. The superior performance of Faure sequences suggests that low-discrepancy sequences are a promising avenue for future research in making ELA-based models more robust and accurate.

Evolutionary algorithms; parameter tuning; landscape analysis.

Введение. Успехи методов машинного обучения (МО) в последние годы оказывают влияние практически на все научные области. Компьютерные науки – не исключение, и одна из ключевых областей применения МО – это выбор и настройка эвристик оптимизации. Как было показано в прошлом исследовании [1], анализ ландшафта приспособленности (FLA) предоставляет мощный инструмент для извлечения знаний о задаче, что позволяет строить модели машинного обучения для рекомендации параметров алгоритмов [2].

В работе [1] была успешно применена нейронная сеть, обученная на признаках ландшафта, для подбора статических параметров генетического алгоритма. Однако этот и другие подобные подходы [3], основанные на обучении с учителем, критически зависят от качества и способа получения этих самых признаков. Формальнее, признак – это отображение $f: X \subseteq R^n \rightarrow R$ в вещественное число. Такой признак может измерять, например, гладкость или мультимодальность ландшафта, наличие на нем “пиков” и “плато”.

На практике многие численные оптимизационные задачи являются задачами “чёрного ящика”, и, соответственно, не имеют явной модели и доступны только через оценку точек $x \in X$. В случаях, когда $f(x)$ задана неявно, значения признаков приходится приближать на основе набора пар $(x, f(x))$. Изучение подобных приближений лежит в основе анализа ландшафта поиска (ELA) [4]. Методы ELA успешно применялись, например, в оптимизации гиперпараметров для конкретных экземпляров задач [1] и в выборе алгоритмов [5]. При использовании ELA для задачи “чёрного ящика” в первую очередь необходимо определить, сколько точек брать и как их генерировать.

При выборе количества образцов для вычисления признаков ELA существует компромисс между точностью оценки и вычислительными затратами. Для быстрых функций рекомендуется использовать порядка $(50-100)n$ образцов, тогда как для ресурсоемких вычислений это число обычно сокращают (в некоторых случаях до $(1-10)n$) [6]. Что касается методов выборки, то здесь доминируют равномерное распределение [7] и латинский гиперкуб (LHS), хотя исследуются и альтернативы. Отдельно стоят траекторные методы [8], которые применяются, если есть возможность выбирать точки для расчёта, что выполняется не всегда.

Проведенное исследование выявило, что аппроксимации признаков, полученные разными методами выборки, не сходятся к единым значениям. Таким образом, модели машинного обучения нельзя использовать на наборе данных, собранном другим методом выборки, а при сравнении задач необходимо убедиться, что различия вызваны свойствами ландшафта, а не методом выборки. Кроме того, в работе показано, что использование последовательностей Фауре позволяет строить более качественные модели по сравнению с общепринятыми методами выборки, такими как равномерная и LHS, что подчеркивает важность исследования влияния свойств выборки на качество обучения.

Постановка задачи. Цель работы – исследовать влияние метода и объема выборки на аппроксимацию признаков анализа ландшафта поиска (ELA) и на последующую точность моделей машинного обучения. Предполагается, что метод выборки вносит в признаки ELA не только случайную ошибку, но и систематическое смещение, что ставит под сомнение объективность их значений и совместимость между разными методами семплирования.

В рамках работы проводится сравнительный анализ методов выборки (Mersenne Twister, LHS, Faure) и объемов данных ($k = n, 10n, 100n$) с целью оценки их влияния на качество регрессионных моделей, предсказывающих параметры тестовой задачи W-model, а также выполняется кросс-валидация моделей между разными методами выборки для оценки зависимости точности и надежности всей процедуры аппроксимации признаков ELA от стратегии и объема исходной выборки.

Регрессия параметров W-model. Для изучения влияния методов выборки на аппроксимацию признаков проанализировано их воздействие на качество модели в задаче регрессии параметров функции W-model. В качестве объектов исследования рассматриваются различные конфигурации параметризованной задачи W-model [9]. Каждая конфигурация задается параметрами: dummy, neutrality, epistasis, ruggedness, а также размерностью n . Для каждой конфигурации W-model используется реализация на основе битовых строк, и целью является предсказание оригинальных параметров W-model по аппроксимированным значениям признаков. Диапазоны параметров W-model и размерности соответствуют настройкам, использованным в прошлом исследовании [1] для переиспользования данных.

Для вычисления признаков каждой функции из полученного набора генерируются k случайных точек $x(1), \dots, x(k)$ и вычисляются соответствующие значения функций $f(x(1)), \dots, f(x(k))$. Полученные пары $(x(i), f(x(i)))_{i=1}^k$ обрабатываются пакетом flasso [6], который высчитывает вектор из 35 признаков. Данная процедура повторяется 2000 раз независимо с использованием одного и того же метода выборки. Таким образом для каждой функции получаются 2000 векторов признаков.

Из этого набора случайным образом отбираются 1000 векторов на функцию для обучения модели регрессора. Ее задача – по новому вектору признаков определить параметры функции W-model. Тестирование проводится на оставшихся векторах, не использованных при обучении, вычислением среднеквадратической ошибки между предсказанными и истинными значениями параметров. Процедура случайного разделения векторов на обучающую и тестовую выборки повторяется 50 раз для повышения достоверности результатов.

Для исследования влияния объема выборки эксперимент проводится при трех значениях k : $k = n, k = 10n, k = 100n$, где n – размерность соответствующей выборке функции W-model.

Конфигурации W-model охватывают широкий спектр свойств ландшафта целевой функции, что делает задачу регрессии параметров нетривиальной и показательной для оценки устойчивости признаков.

Распределения значений признаков. На рис. 1 представлено распределение аппроксимированных значений для конкретного признака, оценивающего степень соответствия задачи квадратичной модели. Результаты показаны для выборок mersenne размером $k = 10n$ и $k = 100n$.

Медианные значения признака (белые линии) демонстрируют существенные межконфигурационные различия между различными параметризациями W-model, что свидетельствует о чувствительности признака к свойствам ландшафта. При этом для каждой отдельной конфигурации наблюдается близкое соответствие медианных оценок, полученных при различных объемах выборки. Однако дисперсия аппроксимации признака значительно снижается с увеличением объема выборки: распределения для $k = 100n$ характеризуются существенно меньшим разбросом по сравнению с $k = 10n$.

Полученные результаты согласуются с данными работы [6] и подтверждают, что увеличение объема выборки повышает устойчивость оценок признаков, хотя межконфигурационные различия сохраняются независимо от объема данных.

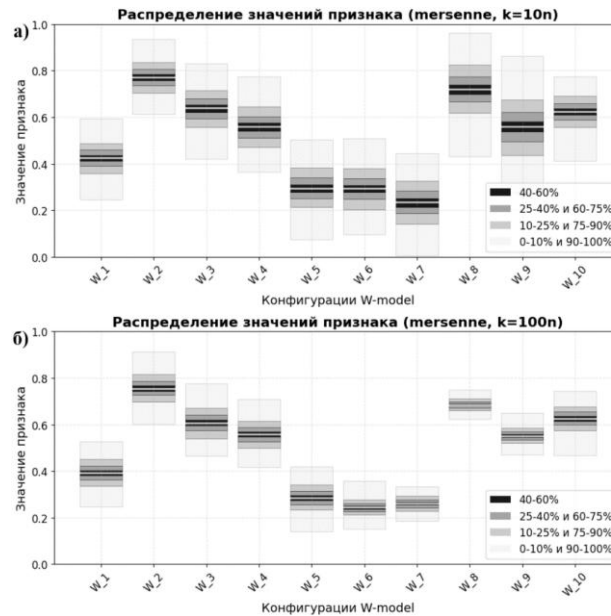


Рис. 1. Распределение значений признака *flacco* на разных конфигурациях *W-model*

Линейная регрессия и градиентный бустинг для задачи регрессии. Все эксперименты выполнены с использованием пакетов *statsmodels* и *XGBoost*. Для задачи регрессии параметров *W-model* применяются: линейная регрессия с L2-регуляризацией (Ridge) [10] и градиентный бустинг (Gradient Boosting) [11]. Поскольку не стоит задачи изучения точности самих моделей, используются стандартные параметры их реализации в соответствующих пакетах.

Использование двух различных регрессоров, линейного с регуляризацией и нелинейного ансамблевого метода, позволяет проверить, наблюдаются ли выявленные эффекты независимо от метода машинного обучения, что помогает исключить специфические артефакты отдельных алгоритмов. Дополнительно была проведена кросс-валидация с 10 запусками модели случайного леса (Random Forest) [12]. Структура основных результатов (в частности, представленных в табл. 1) осталась схожей с полученной при использовании двух основных регрессоров.

Методы выборки. В задачах извлечения признаков наиболее распространенными методами выборки являются равномерное случайное распределение и латинский гиперкуб (LHS). Для исследования зависимости устойчивости аппроксимации признаков от метода выборки, были рассмотрены пять различных методов выборки:

- ♦ *mersenne*: вихрь Мерсенна [13], являющийся стандартным генератором во многих языках программирования и информационных системах. Считается достаточно надежным решением и в данной работе рассматривается в качестве основного генератора для равномерного распределения.

- ♦ *gandu*: линейный конгруэнтный генератор псевдослучайных чисел [14]. Данный генератор считается неудачным решением и не проходит спектральный тест. В работе рассматривается для оценки влияния качества генератора на результаты задачи регрессии параметров *W-model* и, соответственно, аппроксимированные признаки ландшафта.

- ♦ *lhs*: это статистический метод для создания псевдослучайной выборки значений параметров из многомерного распределения [15]. В LHS новые точки выбираются таким образом, чтобы избежать совпадения координат с ранее отобранными точками. Диапазон каждой координаты разбивается на *k* равных интервалов. Из образованной сетки точки отбираются так, чтобы каждая одномерная проекция содержала ровно одну точку в каждом интервале.

♦ *ilhs*: оптимизация генератора LHS [16], использующая жадную эвристику для выбора точек. На каждом шаге из нескольких случайных кандидатов выбирается та точка, которая максимизирует минимальное расстояние до уже выбранных точек.

♦ *faure*: последовательности, предложенные Фауре в [17] являются развитием последовательностей Холтона [18] и известны своим равномерным покрытием пространства [19]. Данный метод основывается на перестановках цифр в разложении чисел по основанию, которым служит наименьшее простое число, не меньшее размерности пространства, что гарантирует отсутствие корреляций между проекциями.

Качество регрессии параметров W-model. На рис. 2 представлено распределение ошибки регрессии, достигнутой для каждого из пяти методов выборки, при условии использования одинаковых методов для построения и оценки регрессионных моделей. Результаты слева соответствуют модели линейной регрессии с L2-регуляризацией, справа – градиентному бустингу.

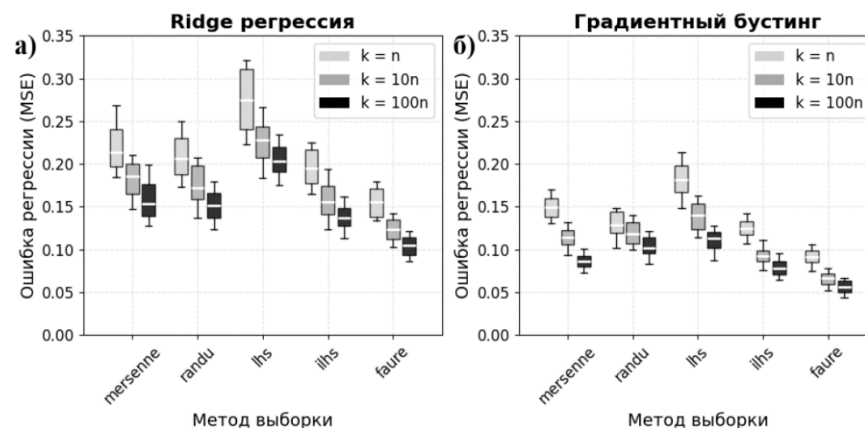


Рис. 2. Ошибка моделей регрессии, обученных на данных различных методов выборки

Как и ожидалось, увеличение объема выборки приводит к улучшению качества регрессии. Можно также заметить, что результаты Ridge регрессии хуже результатов градиентного бустинга. Этот результат закономерен, поскольку линейная модель Ridge, как правило, уступает в точности нелинейному градиентному бустингу на задачах сложной структуры. Однако, сравнение самих моделей между собой выходит за рамки этой работы.

Для регрессора на основе градиентного бустинга четко наблюдается уменьшение разброса ошибки предсказания с ростом объема выборки от $k = n$ к $k = 100n$. Это согласуется с теоретическим снижением дисперсии аппроксимации признаков, обсуждавшимся ранее. В то же время, для линейной регрессии дисперсия ошибки существенно уменьшается при переходе от $k = n$ к $k = 10n$, но затем стабилизируется при дальнейшем увеличении объема данных до $k = 100n$, что может указывать на принципиальное ограничение этой модели в рамках данной задачи.

Не наблюдается существенных различий между двумя генераторами псевдослучайных чисел (*mersenne* и *randu*) по их влиянию на ошибку регрессии. В случае с LHS, напротив, базовый метод (*lhs*) показывает значительно большую ошибку предсказания по сравнению с его улучшенной версией (*ilhs*) на всех объемах выборки.

Для метода выборки на основе последовательностей Фауре (*faure*) в большинстве рассмотренных комбинаций (модель регрессии, объем выборки) наблюдается наименьшая медианная ошибка регрессии. Графики для этого метода, особенно при объемах выборки $k = n$ и $k = 10n$, хорошо отделены от других методов, что визуально свидетельствует о его эффективности. Статистические тесты подтверждают значимость этих различий практически во всех случаях.

Результаты проведенного экспериментального исследования свидетельствуют о существенном влиянии метода выборки на качество аппроксимации признаков и последующую точность моделей машинного обучения. Наблюдаемое устойчивое превосходство детерминированных последовательностей Фауре (faure) над общепринятыми методами (равномерное распределение и LHS) ставит под сомнение целесообразность их использования по умолчанию в задачах анализа ландшафта поиска (ELA). Однако, применение последовательностей Фауре в машинном обучении требует больше вычислительных ресурсов для их генерации, по сравнению с общепринятыми методами.

Статистический анализ не выявил значимых преимуществ методов, основанных на псевдослучайных числах, над базовым LHS – наблюдаемые различия были статистически незначимыми. В то же время метод на основе последовательностей Фауре продемонстрировал статистически значимое снижение ошибки регрессии по всему диапазону исследуемых объемов выборки (от $k = n$ до $k = 100n$) и для всех протестированных алгоритмов машинного обучения. Устойчивость данного результата, подтвержденная различными регрессорами, указывает на то, что низкое отклонение квазислучайных последовательностей обеспечивает более точную и стабильную аппроксимацию признаков ландшафта.

Влияние метода выборки. Из предыдущего анализа следует, что вариативность ошибки регрессии при изменении объема выборки в значительной степени объясняется дисперсией аппроксимированных значений признаков. Вместе с тем сравнительный анализ различных методов демонстрирует, что одна лишь дисперсия не может служить основным фактором, объясняющим наблюдаемые между ними различия. Следовательно, на эффективность методов влияют иные механизмы, требующие дополнительного изучения.

На рис. 3 представлено распределение аппроксимированных значений для двух из 35 рассматриваемых признаков. Анализ визуализации выявил существенную зависимость оценок признаков от метода выборки, что противопоставляет традиционному представлению об их абсолютном характере. Для признака, представленного на левом графике, распределения не только обладают различными медианами и средними значениями, но и являются полностью разделимыми. Данная закономерность устойчиво проявляется при различных объемах выборки.

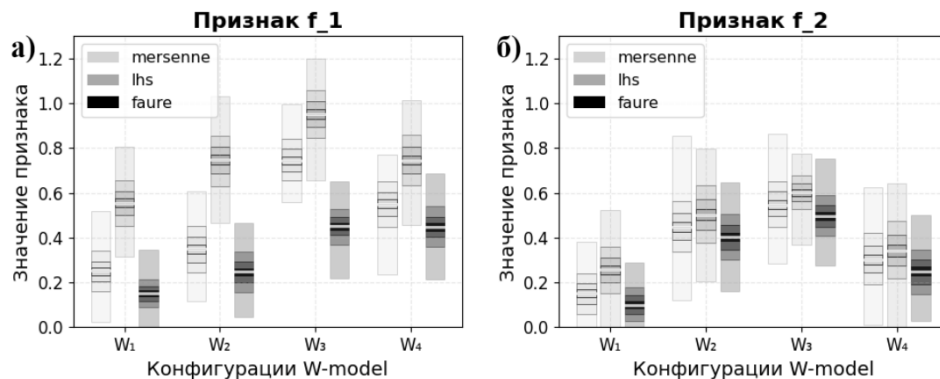


Рис. 3. Распределения двух признаков фласко для разных конфигураций W-model и методов выборки

Полученные результаты свидетельствуют против предположения о возможности прямого переноса оценок признаков между различными методами выборки. В качестве иллюстрации рассмотрим две конфигурации W-model: W_1 ($n = 256$, $dumtuy = 0.9$, $neutrality = 2$, $epistasis = 2$, $ruggedness = 5$) и W_2 ($n = 256$, $dumtuy = 0.8$, $neutrality = 6$, $epistasis = 3$, $ruggedness = 10$). Анализ показывает, что методы случайной выборки (mersenne) и квазислучайных последовательностей Фауре (faure) дают согласованные оценки признаков для каждого ландшафта в отдельности. В то же время аппроксимации на основе латинского гиперкуба (LHS) демонстрируют статистически значимое за-

вышение значений определенных признаков для W_1 относительно W_2 . Следовательно, корректная интерпретация значений признаков требует обязательного учета использованного метода выборки, поскольку наблюдаемые межзадачные различия могут быть обусловлены не фундаментальными свойствами ландшафта, а артефактами процедуры семплирования.

Зависимость аппроксимации признаков от выборки. С целью исследования зависимости устойчивости аппроксимаций признаков от методов выборки был реализован специализированный эксперимент в контексте задачи регрессии. В рамках данной схемы регрессоры обучались на векторах признаков, вычисленных по выборкам, сгенерированным одним методом, а их обобщающая способность оценивалась на тестовых наборах, полученных с применением альтернативных методов выборки. Используемый набор признаков и общая методология подготовки данных соответствуют подходу, изложенному в предыдущих разделах, с единственным изменением, заключающимся в кросс-методологическом разделении исходных данных на обучающую и тестовую выборки.

В табл. 1 представлены данные, отражающие медианную ошибку регрессионных моделей для всех 25 комбинаций методов выборки, использованных при формировании обучающих и тестовых наборов данных. Таблица демонстрирует результаты для трех объемов выборки: $k = n$, $k = 10n$ и $k = 100n$. Элементы главной диагонали, соответствующие случаю совпадения методов на этапах обучения и тестирования, были проанализированы ранее. Левая часть таблицы содержит результаты для Ridge-регрессии, правая – для градиентного бустинга.

Таблица 1

Медианные ошибки регрессии параметров W-model

a) Ridge регрессия (k = n)					
	mers	randu	lhs	ilhs	faure
mers	0.245	0.312	0.298	0.285	0.401
randu	0.315	0.248	0.305	0.292	0.408
lhs	0.301	0.308	0.285	0.275	0.395
ilhs	0.288	0.295	0.278	0.195	0.382
faure	0.405	0.412	0.398	0.385	0.152
b) Gradient перрессия (k = n)					
	mers	randu	lhs	ilhs	faure
mers	0.145	0.212	0.198	0.185	0.301
randu	0.215	0.148	0.205	0.192	0.308
lhs	0.201	0.208	0.185	0.175	0.295
ilhs	0.188	0.195	0.178	0.125	0.282
faure	0.305	0.312	0.298	0.285	0.092
в) Ridge регрессия (k = 10n)					
	mers	randu	lhs	ilhs	faure
mers	0.198	0.245	0.232	0.218	0.335
randu	0.248	0.201	0.238	0.225	0.342
lhs	0.235	0.242	0.235	0.208	0.328
ilhs	0.221	0.228	0.215	0.165	0.315
faure	0.338	0.345	0.331	0.318	0.128
г) Gradient перрессия (k = 10n)					
	mers	randu	lhs	ilhs	faure
mers	0.098	0.145	0.132	0.118	0.235
randu	0.148	0.101	0.138	0.125	0.242
lhs	0.135	0.142	0.135	0.108	0.228
ilhs	0.121	0.128	0.115	0.095	0.215
faure	0.238	0.245	0.231	0.218	0.068
д) Ridge регрессия (k = 100n)					
	mers	randu	lhs	ilhs	faure
mers	0.175	0.212	0.198	0.185	0.285
randu	0.215	0.178	0.205	0.192	0.292
lhs	0.202	0.208	0.210	0.178	0.278
ilhs	0.188	0.195	0.185	0.148	0.265
faure	0.288	0.295	0.281	0.268	0.115
е) Gradient перрессия (k = 100n)					
	mers	randu	lhs	ilhs	faure
mers	0.075	0.112	0.098	0.085	0.185
randu	0.115	0.078	0.105	0.092	0.192
lhs	0.102	0.108	0.110	0.078	0.178
ilhs	0.088	0.095	0.085	0.078	0.165
faure	0.188	0.195	0.181	0.168	0.055

Для объемов выборки $k = 10n$ и $k = 100n$ наименьшая, или близкая к наименьшей, ошибка регрессионной модели достигается при условии совпадения методов формирования выборок на этапах обучения и тестирования. Указанная закономерность является статистически устойчивой и проявляется независимо от типа используемого регрессора (Ridge-регрессия или градиентный бустинг), что визуально подтверждается наименьшими значениями ошибки на элементах главной диагонали матриц, представленных в табл. 1.

Для минимального объема выборки $k = n$ регистрируется значительный разброс значений ошибки регрессии, что свидетельствует о низкой устойчивости аппроксимации признаков. При этом, несмотря на высокую сложность задачи прогнозирования парамет-

ров W-model, обусловленную многомерностью пространства признаков и разнообразием конфигураций ландшафта, абсолютное значение ошибки предсказания оказывается приемлемым, особенно в случае применения нелинейного метода градиентного бустинга.

Анализ случаев с объемом выборки $k \geq 10n$ показывает значительное уменьшение ошибки регрессии, что согласуется с теоретическим снижением дисперсии аппроксимации признаков. Ключевым фактором, определяющим эффективность моделей, является степень согласованности методов выборки на этапах обучения и тестирования. Наибольшая ошибка наблюдается при принципиальном различии этих методов (например, *mersenne-lhs*, *lhs-faure*, *randu-ilhs*). Промежуточные значения ошибки достигаются в случае использования схожих методов *lhs-ilhs*. При этом минимальная ошибка регрессии обеспечивается только при полном совпадении методов генерации выборок для обучающей и тестовой выборок, что подтверждается результатами для всех пар типа *mersenne-mersenne*, *faure-faure* и т.д. Данная закономерность имеет систематический характер и наблюдается для обоих регрессоров.

Анализ перекрестного тестирования выявил специфическую особенность метода Фауре. В то время как регрессоры, обученные на признаках, аппроксимированных по другим методам, демонстрируют сопоставимый уровень ошибки при тестировании на данных иного типа, модель, обученная на выборках Фауре, показывает статистически значимо более высокую ошибку при валидации на любом из альтернативных методов. Данный эффект наблюдается для обоих регрессоров, однако является более выраженным для градиентного бустинга по сравнению с моделью линейной регрессии.

Результаты детального анализа распределений аппроксимированных признаков позволяют выявить кластеризацию методов выборки по степени схожести генерируемых ими оценок. Наблюдается высокая степень согласованности между методами *lhs* и *ilhs*, значения признаков для которых практически не отличаются. В то же время, методы *faure* и *randu* формируют обособленные группы, существенно отклоняясь не только от пары *lhs/ilhs*, но и от стандартной случайной выборки *mersenne*. Количественная оценка данного эффекта показывает, что почти для половины всех признаков на всех конфигурациях W-model при $k = 100n$, медианное значение признаков, полученных с использованием *faure* или *randu* статистически значимо отличается от медиан, полученных другими методами.

Таким образом, результаты кросс-методологической валидации однозначно демонстрируют, что аппроксимированные признаки не являются инвариантными относительно метода выборки. Их значения и, как следствие, информативность для моделей машинного обучения критически зависят от метода выборки. Это означает, что признаки, полученные разными методами, несут в себе не только информацию о свойствах ландшафта, но и специфический «отпечаток» способа их генерации. Следовательно, для построения корректных прогнозных моделей в рамках ELA необходимо строгое соответствие между методами выборки, использованными на этапах обучения и применения.

Ошибки моделей регрессии. Для углубленного анализа характера ошибок, допускаемых регрессионными моделями, был проведен детальный разбор предсказаний. Результаты анализа показывают, что для большинства конфигураций W-model медианная ошибка предсказания отдельных параметров является достаточно низкой. Однако наблюдаются систематические смещения для определенных диапазонов значений параметров. Так, для конфигураций с высокими значениями параметра *ruggedness* (например, > 8) модель имеет тенденцию к устойчивому занижению предсказаний. Аналогичным образом, для параметра *dumtmy* в области значений около 0.5 часто наблюдается бимодальное распределение ошибок, где значительная часть предсказаний образует кластер вокруг истинного значения, а другая – смещенный кластер в области более высоких значений.

Эти систематические ошибки проявляются согласованно для множественных повторений эксперимента на одних и тех же конфигурациях W-model. Это указывает на то, что они обусловлены не случайным шумом, а фундаментальными особенностями взаимодействия между конкретными свойствами ландшафта и способом их отражения в аппроксимированных признаках при данном методе выборки.

Подобный анализ ошибок полезен для выявления слабых мест как в наборе признаков, так и в методах их оценки. Устойчивое смещение предсказаний для конкретных параметров может указывать на необходимость введения новых, более специализированных признаков, чувствительных к соответствующим свойствам ландшафта [20]. Кроме того, различия в структуре ошибок между разными методами выборки подчеркивают, что выбор метода семплирования влияет не только на общую точность модели, но и на то, какие именно конфигурации будут распознаваться хуже или лучше.

Выводы. В работе проведено исследование влияния метода и объема выборки на точность аппроксимации признаков в рамках анализа ландшафта поиска (ELA) и на последующее качество моделей машинного обучения.

Установлено, что аппроксимированные значения признаков ELA не являются абсолютными характеристиками ландшафта и существенно зависят от примененного метода выборки. Различные методы выборки приводят к статистически значимо различным оценкам одних и тех же признаков даже при больших объемах данных. Следствием этого является критическая важность использования единого метода генерации выборок на этапах обучения и применения моделей машинного обучения для обеспечения их корректной работы.

Обнаружено, что модели, обученные на признаках, полученных с помощью квазислучайных последовательностей Фауре, демонстрируют меньшую ошибку в задаче регрессии параметров W-model по сравнению с моделями, использующими признаки на основе стандартных методов (равномерное распределение, латинский гиперкуб). Это указывает на перспективность дальнейшего исследования детерминированных последовательностей с низким отклонением для задач ELA.

Проведенная кросс-методологическая валидация подтвердила, что наилучшее качество прогноза достигается только при строгом совпадении метода выборки для обучающего и тестового наборов. Наблюдаемые систематические ошибки предсказания для отдельных диапазонов параметров W-model подчеркивают сложность взаимодействия между свойствами ландшафта и методом их оценки.

Полученные результаты свидетельствуют о необходимости учета специфики методов выборки при интерпретации признаков ELA и разработке прогнозных моделей. Для дальнейшего повышения надежности методов ELA перспективными направлениями являются исследование функциональной связи между методами выборки и значениями признаков, а также разработка методов коррекции полученных оценок.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Пикалов М.В., Письмеров А.М. Настройка параметров генетического алгоритма при помощи анализа ландшафта функции приспособленности и машинного обучения // Известия ЮФУ. Технические науки. – 2024. – № 2 (238). – С. 221-228.
2. Lobo F.J., Lima C.F., Michalewicz Z. (ed.). Parameter setting in evolutionary algorithms. – Springer Science & Business Media, 2007. – Т. 54.
3. Mersmann O. et al. Exploratory landscape analysis // Proceedings of the 13th annual conference on Genetic and evolutionary computation. – 2011. – P. 829-836.
4. Ochoa G., Malan K. Recent advances in fitness landscape analysis // Proceedings of the genetic and evolutionary computation conference companion. – 2019. – P. 1077-1094.
5. Kerschke P., Trautmann H. Automated algorithm selection on continuous black-box problems by combining exploratory landscape analysis and machine learning.
6. Kerschke P., Trautmann H. The R-Package FLACCO for exploratory landscape analysis with applications to multi-objective optimization problems // 2016 IEEE Congress on Evolutionary Computation (CEC). – IEEE, 2016. – P. 5262-5269.
7. Mersmann O., Preuss M., Trautmann H. Benchmarking evolutionary algorithms: Towards exploratory landscape analysis. – 2010.
8. Huang C., Li Y., Yao X. A survey of automatic parameter tuning methods for metaheuristics // IEEE transactions on evolutionary computation. – 2019. – Vol. 24, No. 2. – P. 201-216.
9. Weise T., Wu Z. Difficult features of combinatorial optimization problems and the tunable w-model benchmark problem for simulating them // Proceedings of the Genetic and Evolutionary Computation Conference Companion. – 2018. – P. 1769-1776.

10. McDonald G.C. Ridge regression // *Wiley Interdisciplinary Reviews: Computational Statistics*. – 2009. – Vol. 1, No. 1. – P. 93-100.
11. Friedman J.H. Greedy function approximation: a gradient boosting machine // *Annals of statistics*. – 2001. – P. 1189-1232.
12. Breiman L. Random forests // *Machine learning*. – 2001. – Vol. 45, No. 1. – P. 5-32.
13. Matsumoto M., Nishimura T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator // *ACM Transactions on Modeling and Computer Simulation (TOMACS)*. – 1998. – Vol. 8, No. 1. – P. 3-30.
14. Marsaglia G. Random numbers fall mainly in the planes // *Proceedings of the National Academy of sciences*. – 1968. – Vol. 61, No. 1. – P. 25-28.
15. Helton J.C., Davis F.J. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems // *Reliability Engineering & System Safety*. – 2003. – Vol. 81, No. 1. – P. 23-69.
16. Huntington D.E., Lyrantzis C.S. Improvements to and limitations of Latin hypercube sampling // *Probabilistic engineering mechanics*. – 1998. – Vol. 13, No. 4. – P. 245-253.
17. Faure H., Lemieux C. Generalized Halton sequences in 2008: A comparative study // *ACM Transactions on Modeling and Computer Simulation (TOMACS)*. – 2009. – Vol. 19, No. 4. – P. 1-31.
18. Wang X., Hickernell F.J. Randomized halton sequences // *Mathematical and Computer Modelling*. – 2000. – Vol. 32, No. 7-8. – P. 887-899.
19. Faure H., Lemieux C. Generalized Halton sequences in 2008: A comparative study // *ACM Transactions on Modeling and Computer Simulation (TOMACS)*. – 2009. – Vol. 19, No. 4. – P. 1-31.
20. Pitzer E., Affenzeller M. A comprehensive survey on fitness landscape analysis // *Recent advances in intelligent engineering systems*. – 2012. – P. 161-191.

REFERENCES

1. Pikalov M.V., Pis'merov A.M. Nastroyka parametrov geneticheskogo algoritma pri pomoshchi analiza landshafta funktsii prispособlennosti i mashinnogo obucheniya [Tuning genetic algorithm parameters using fitness landscape analysis and machine learning], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2024, No. 2 (238), pp. 221-228.
2. Lobo F.J., Lima C.F., Michalewicz Z. (ed.). Parameter setting in evolutionary algorithms. Springer Science & Business Media, 2007, Vol. 54.
3. Mersmann O. et al. Exploratory landscape analysis, *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, 2011, pp. 829-836.
4. Ochoa G., Malan K. Recent advances in fitness landscape analysis, *Proceedings of the genetic and evolutionary computation conference companion*, 2019, pp. 1077-1094.
5. Kerschke P., Trautmann H. Automated algorithm selection on continuous black-box problems by combining exploratory landscape analysis and machine learning.
6. Kerschke P., Trautmann H. The R-Package FLACCO for exploratory landscape analysis with applications to multi-objective optimization problems, *2016 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2016, pp. 5262-5269.
7. Mersmann O., Preuss M., Trautmann H. Benchmarking evolutionary algorithms: Towards exploratory landscape analysis, 2010.
8. Huang C., Li Y., Yao X. A survey of automatic parameter tuning methods for metaheuristics, *IEEE transactions on evolutionary computation*, 2019, Vol. 24, No. 2, pp. 201-216.
9. Weise T., Wu Z. Difficult features of combinatorial optimization problems and the tunable w-model benchmark problem for simulating them, *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 2018, pp. 1769-1776.
10. McDonald G.C. Ridge regression, *Wiley Interdisciplinary Reviews: Computational Statistics*, 2009, Vol. 1, No. 1, pp. 93-100.
11. Friedman J.H. Greedy function approximation: a gradient boosting machine, *Annals of statistics*, 2001, pp. 1189-1232.
12. Breiman L. Random forests, *Machine learning*, 2001, Vol. 45, No. 1, pp. 5-32.
13. Matsumoto M., Nishimura T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 1998, Vol. 8, No. 1, pp. 3-30.
14. Marsaglia G. Random numbers fall mainly in the planes, *Proceedings of the National Academy of sciences*, 1968, Vol. 61, No. 1, pp. 25-28.
15. Helton J.C., Davis F.J. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, *Reliability Engineering & System Safety*, 2003, Vol. 81, No. 1, pp. 23-69.

16. Huntington D.E., Lyrantzis C.S. Improvements to and limitations of Latin hypercube sampling, *Probabilistic engineering mechanics*, 1998, Vol. 13, No. 4, pp. 245-253.
17. Faure H., Lemieux C. Generalized Halton sequences in 2008: A comparative study, *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 2009, Vol. 19, No. 4, pp. 1-31.
18. Wang X., Hickernell F.J. Randomized halton sequences, *Mathematical and Computer Modelling*, 2000, Vol. 32, No. 7-8, pp. 887-899.
19. Faure H., Lemieux C. Generalized Halton sequences in 2008: A comparative study, *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 2009, Vol. 19, No. 4, pp. 1-31.
20. Pitzer E., Affenzeller M. A comprehensive survey on fitness landscape analysis, *Recent advances in intelligent engineering systems*, 2012, pp. 161-191.

Пикалов Максим Вадимович – Национальный исследовательский университет ИТМО; e-mail: pikmaksim@gmail.com; г. Санкт-Петербург, Россия; тел.: 88126070283; аспирант.

Pikalov Maxim Vadimovich– ITMO University; e-mail: pikmaksim@gmail.com; St. Petersburg, Russia; phone: +78126070283; graduate student.

УДК 004.032.26

DOI 10.18522/2311-3103-2026-1-52-64

М.О. Доброхвалов, А.Ю. Филатов, Е.А. Чегодаева

НОВАЯ МЕТРИКА ВОСПРОИЗВОДИМОСТИ ДЛЯ СРАВНЕНИЯ КЛАССИФИКАТОРОВ ВРЕМЕННЫХ РЯДОВ

Воспроизводимость результатов экспериментов является критическим аспектом современного машинного обучения, однако выбор случайного инициализирующего сета существенно влияет на итоговое качество моделей, что создает проблему корректного сравнения различных архитектур и методов. Цель исследования заключалась в оценке влияния выбора случайного сета на результаты классификации временных рядов сверточными нейросетями и в разработке корректного способа сравнения моделей. Задачи включали измерение разброса метрик при множественных повторных запусках в рамках варьирующихся инициализаций, проверку нормальности распределений, введение метрики воспроизводимости RM и подбор ее параметра λ , а также проверку переносимости подхода на альтернативных архитектурах. Проведены эксперименты с двумя одномерными архитектурами (FCN, ResNet) сверточных нейронных сетей на семи открытых наборах данных временных рядов разной природы. Для каждой пары модель–датасет выполнено по 55 независимых запусков с фиксацией источников случайности и идентичными настройками обучения в PyTorch. Статистический анализ включал критерии Шапиро Уилка и Андерсона Дарлинга. Показано, что распределения аккуратности чаще всего не соответствуют нормальному закону, поэтому интервальные оценки, основанные на нормальности, некорректны. Варьирование сета приводит к различиям аккуратности до 12 процентных пунктов, причем величина разброса зависит от датасета и архитектуры. Предложенная метрика воспроизводимости (RM), штрафующая за дисперсию, при малом числе запусков на различных инициализирующих значениях приближает нижнюю наблюдаемую границу, при большом числе стремится к среднему. Предложенная RM позволяет сравнивать модели с учетом случайных “удачных” и “неудачных” инициализаций и ранжировать модели по устойчивости, задает стандарт отчетности, повышающий воспроизводимость экспериментов и надежность выводов. Эмпирическая проверка на архитектуре DenseNet подтвердила, что RM адекватно реагирует как на стабильные, так и на нестабильные наборы. Методика легко переносится на новые датасеты и архитектуры. Предложенная метрика может быть использована для стандартизации отчетности и повышения воспроизводимости исследований.

Классификация временных рядов; сверточные нейронные сети; случайный сет; чувствительность к инициализации; воспроизводимость экспериментов.

M.O. Dobrokhvalov, A.Yu. Filatov, E.A. Chegodaeva

A NEW REPRODUCIBILITY METRIC FOR COMPARING TIME SERIES CLASSIFIERS

Experimental reproducibility constitutes a critical cornerstone of modern machine learning research, yet random initialization seed selection substantially influences final model performance, creating challenges for principled comparison of different architectures and methods. Random seed effects on convolutional time series classifiers were quantified, and a principled comparison criterion was established. Two 1D architectures, FCN and ResNet, were trained on seven public datasets containing different data. 55 independent runs for each combination of model and dataset were performed under controlled pseudorandomness in Python, NumPy, and PyTorch. Deterministic backends were enabled, and identical hyperparameters were used across runs. Normality of seed-wise accuracy distributions was assessed with the Shapiro–Wilk and Anderson–Darling tests. Accuracy variability attributable to seed choice reached up to 12 percentage points in some settings, with magnitude dependent on dataset and architecture. The distributions were found to be non-normal in most cases, indicating that confidence intervals predicated on normality are unreliable. To enable fair comparison across runs, a reproducibility meta-metric, RM, was introduced that subtracts a dispersion penalty from the mean and depends on the number of runs and a tunable coefficient λ . RM was shown to lie between the empirical minimum and the mean, to approach the lower bound for small sample sizes, and to converge toward the mean as the number of runs increases. Portability of the approach was examined on an additional architecture, DenseNet, confirming expected behavior. Practical value is provided by RM metric rankings reflect both performance and stability. In this way, reproducibility and the credibility of empirical conclusions are strengthened.

Time series classification; convolutional neural networks; random seed; initialization sensitivity; experimental reproducibility.

Введение. Машинное обучение продолжает активно развиваться, в частности, нейронные сети. До начала процесса обучения веса сети инициализируются определенным образом. Инициализация весов сверточной сети влияет на конечное состояние, к которому сходится модель. Существуют различные методы инициализации весов [1].

Однако, даже в рамках одного метода инициализации веса могут иметь различные значения, которые зависят от состояния генератора случайных чисел. Это может вносить различия в результаты обучения. В работе [2] рассматривается влияние инициализирующего сида на результаты классификации с использованием разновидностей модели VGG на наборе данных CIFAR-10 [3]. Авторами рассматривались три способа выбора значения для сида: (i) последовательные значения [0, 19], (ii) получение псевдослучайного значения с помощью библиотеки `numpy`¹ с инициализацией состояния с помощью последовательного выбора значений [0, 19], (iii) использование стандартной инициализации без фиксирования состояния. Авторы демонстрируют, что разность между минимальным и максимальным значением аккуратности (ассигасы) может составлять более 3, при среднем значении 90,584 и стандартном отклонении 0,216. В [4] рассматривают вероятностную модель скрытых состояний для описания динамики обучения нейросетей с использованием 40 сидов.

В работе [5] также рассматривается результат влияния инициализирующего сида на результаты предсказания. Автор рассматривает 3 различных архитектуры/инициализации на 2 датасетах: CIFAR-10 [6] и ImageNet [7]. Автор демонстрирует, что в рамках датасета CIFAR-10 разность между минимальным и максимальным значениями метрики аккуратности может достигать 1,82 при среднем значении 90,02 и стандартном отклонении 0,23. При использовании датасета ImageNet стандартное отклонение составляет около 0,1, а разность между минимальным и максимальным значениями составляет около 0,5 при среднем значении около 76.

В работе [8] предлагается механизм агрегации с вниманием для повышения точности сверточных сетей, при этом все эксперименты проводились с фиксированным сидом 0. В статье [9] для всех экспериментов в области генеративного искусственного интеллекта фиксировался сид 42.

¹ <https://numpy.org/doc/stable/>.

На основании [10] в ряде работ применяли множество сидов для количественной оценки устойчивости и статистической надёжности предлагаемых методов. В [11] использовали 11 сидов при решении задачи одновременного определения типов клеток и сегментации тканей, в [12] провели эксперименты с 5 сидами в задаче детектирования объектов, в [13] рассмотрели 100 сидов при оптимизации процедуры обучения ResNet для решения задачи классификации, в [14] оценили влияние 50 сидов на точность медицинской сегментации, в [15] применили 10 сидов для отбора наиболее качественных результатов преобразования текста в изображение. Также в работах [16] авторы указывают, что фиксируют сид для исключения случайного выбора “удачного сида”.

При сравнении исследований обычно используются результаты, указанные другими авторами. При этом повторить эксперимент часто не представляется возможным из-за недостаточного описания постановки.

В данной работе рассмотрено влияние инициализирующего сида на результаты классификации двух сверточных сетей, обрабатывающих временные ряды, взятые из 7 открытых наборов данных.

Формализация задачи. Пусть задан набор данных $D = \{(x_i, y_i)\}_{i=1}^N$, где $x_i \in R^T$ – временной ряд длины T , а $y_i \in \{1, 2, \dots, K\}$ – метка класса. Требуется обучить классификатор $f_\theta: R^T \rightarrow \{1, 2, \dots, K\}$ с параметрами θ^0 , минимизирующий функцию потерь. Итоговые параметры θ^* зависят от случайного сида $s \in N$, определяющего начальную инициализацию весов $\theta^0(s)$. Для фиксированного датасета и архитектуры получаем множество значений метрики качества $\{m(s_1), m(s_2), \dots, m(s_n)\}$, где $m(s_j)$ – значение метрики качества модели, обученной после инициализации весов s_j . Требуется оценить статистические значения набора метрик качества.

Модели и данные. В дальнейших исследованиях использовалось архитектуры сверточных нейронных сетей: Fully Convolutional Network (FCN) [17], ResNet [18].

Кодировщик архитектуры FCN, основан на более ранних сверточных нейронных сетях для классификации изображений, таких как VGG-16. Ключевой модификацией является замена полностью связанных слоёв на сверточные слои, что позволяет сети работать с изображениями произвольного размера и сохранять пространственные характеристики данных. Энкодер, использующийся в данной работе как сеть для классификации, представляет собой серию сверточных и слоёв пулинга, которые постепенно уменьшают пространственное разрешение, извлекая высокоуровневые признаки из входного изображения. На выходе кодировщика формируются семантически богатые, но низкоразмерные карты признаков, которые затем используются для последующего восстановления пиксельных классификаций в декодере. Также в оригинальной работе рассматривается сеть для обработки изображений. В данной работе она была адаптирована для обработки одномерных сигналов, путём замены двумерных слоев на одномерные.

Также рассматривается адаптированная для решения задачи классификации одномерных временных рядов архитектура сети ResNet. Представленная модификация архитектуры сохраняет ключевые элементы исходной сети [19], при этом учитывает специфику временных данных. Кодировщик сети состоит из серии последовательно соединённых остаточных (residual) блоков, каждый из которых включает три свёрточных слоя с фильтрами размеров 1×1 , 3×3 и снова 1×1 . Особое внимание уделяется механизму остаточных соединений (skip-connections), которые обеспечивают прямую передачу информации между входом и выходом блока, который позволяет уменьшить эффект исчезновения градиентов и упрощает процесс обучения более глубоких сетей. После последовательности резидуальных блоков используется слой глобального среднего объединения (global average pooling), который агрегирует информацию вдоль временного измерения. Архитектура завершается полностью связанным выходным слоем. Данная архитектура также дополнительно адаптирована в рамках данной работы для работы с одномерными данными.

В рамках данного исследования было использовано 7 наборов данных².

Датасет *Adias* состоит из временных рядов длиной 176 отсчётов, представляющих контуры изображений объектов, сгруппированных в 37 классов. Он содержит 390 обучающих и 391 тестовый экземпляр. Задача — классификация форм по контурам. *ECG5000* основан на 20-часовой ЭКГ пациента с сердечной недостаточностью. Из записи были выделены и интерполированы сердечные циклы, после чего случайным образом отобрано 5000 циклов, разбитых на 5 классов. Датасет *Fish* включает ряды длиной 463 отсчёта, характеризующие 7 видов рыб (по 175 экземпляров в обучающем и тестовом наборах). *Lightning2* содержит 2 класса спектрограмм молний длиной 637 отсчётов, всего 60 обучающих и 61 тестовых экземпляров. *Lightning7* усложнён за счёт 7 классов и более коротких рядов (319 отсчётов), содержит 70 обучающих и 73 тестовых экземпляра. *SonyAIBORobotSurface1* и *SonyAIBORobotSurface2* содержат сигналы акселерометров робота Sony AIBO, длиной 70 и 65 отсчётов соответственно, с двумя классами. Тренировочных экземпляров 20 и 27, тестовых 601 и 953 соответственно.

Эксперимент

Постановка.

Фреймворк *PyTorch*³, активно применяемый в современных исследованиях, предоставляет инструменты для улучшения воспроизводимости результатов. Основным методом, позволяющим уменьшить неопределённость результатов экспериментов, является фиксация случайного сета (инициализирующего состояния генератора псевдослучайных чисел). Однако, помимо фиксации сета в самом *PyTorch*, необходимо также фиксировать сиды стандартной библиотеки *Python* и библиотеки работы с массивами *NumPy*, поскольку эти библиотеки также генерируют случайные числа, используемые в вычислениях. Кроме того, важным шагом для достижения стабильной воспроизводимости является активация режима детерминированных операций *PyTorch* и отключение бенчмаркинга, который в противном случае может автоматически выбирать наиболее быстрые алгоритмы для операций, вводя дополнительную неопределённость (листинг 1).

Листинг 1

Фиксирование сета

```
random.seed(seed)
np.random.seed(seed)
torch.manual_seed(seed)
torch.backends.cudnn.deterministic = True
torch.backends.cudnn.benchmark = False
```

Также для подачи данных в едином порядке, сид генератора загрузчика данных дополнительно фиксировался значением 2025.

В качестве функции потерь использовалась перекрестная энтропия (*CrossEntropyLoss*), оптимизатор — *Adam* (шаг — 0.001). Для обучения использовалось 500 эпох. Все эксперименты проводились с использованием графического процессора *NVIDIA AD106M (GeForce RTX 4070 Max-Q, 8188 MiB)*.

Результаты эксперимента.

В табл. 1 представлены статистические характеристики аккуратности моделей, решающих задачу классификации, обученных на датасетах с временными рядами. Для исследования взяты простые и хорошо изученные архитектуры. Рассмотрено 55 сидов [1971, 2025].

² <https://www.timeseriesclassification.com/dataset.php>.

³ <https://docs.pytorch.org/docs/stable/index.html>.

Для сети FCN наиболее устойчивые результаты наблюдаются на Fish и SonyAIBORobotSurface1. В первом случае стандартное отклонение составило 0,005 при размахе 0,023, во втором 0,007 при размахе 0,028. Близкий уровень стабильности показан на ECG5000, стандартное отклонение 0,008 при размахе 0,036. Наиболее нестабильное поведение FCN демонстрирует на Lightning7, где стандартное отклонение достигает 0,024, а размах 0,110. Повышенную вариативность видно и на Lightning2, стандартное отклонение 0,015 при размахе 0,082.

Для сети ResNet отклонение и размах также зависят от датасета. Наименьшая вариативность наблюдается на Fish и SonyAIBORobotSurface2, где стандартные отклонения составили 0,005 и 0,005 при размахе 0,017 и 0,023. Сходный уровень стабильности получен на SonyAIBORobotSurface1, стандартное отклонение 0,008 при размахе 0,037. Наибольшая нестабильность зафиксирована на Lightning2 и ECG5000, где стандартное отклонение достигло 0,032 и 0,026 соответственно, а размах 0,115 и 0,123.

Таким образом, для обеих архитектур наибольшую устойчивость обеспечивают Fish, SonyAIBORobotSurface1 и SonyAIBORobotSurface2, тогда как Lightning2 и ECG5000 характеризуются высокой вариативностью результатов. Для FCN к этой группе нестабильных наборов добавляется Lightning7. Стандартное отклонение и размах значений точности остаются ключевыми параметрами при оценке надежности моделей на различных задачах.

Таблица 1

Характеристики аккуратности моделей, обученных на разных сетах

Сеть	Датасет	Средняя	Стандартное отклонение	Минимальная	Максимальная	Размах
FCN	Adiac	0,808	0,015	0,772	0,847	0,074
	ECG5000	0,924	0,008	0,899	0,935	0,036
	Fish	0,969	0,005	0,954	0,977	0,023
	Lightning2	0,696	0,015	0,656	0,738	0,082
	Lightning7	0,772	0,024	0,712	0,822	0,110
	SonyAIBORobotSurface1	0,957	0,007	0,943	0,972	0,028
	SonyAIBORobotSurface2	0,958	0,009	0,933	0,979	0,046
ResNet	Adiac	0,783	0,013	0,757	0,806	0,049
	ECG5000	0,884	0,026	0,798	0,922	0,123
	Fish	0,979	0,005	0,971	0,989	0,017
	Lightning2	0,712	0,032	0,656	0,771	0,115
	Lightning7	0,788	0,024	0,740	0,822	0,082
	SonyAIBORobotSurface1	0,977	0,008	0,957	0,993	0,037
	SonyAIBORobotSurface2	0,970	0,005	0,957	0,980	0,023

Таким образом, исследования, утверждающие, что значение метрики превышает аналоги на десятки или сотни доли акkuratности, в некоторых случаях могут являться примером удачной инициализации весов сети.

Нормальность распределения метрик.

На рис. 1 представлено распределение акkuratности классификации моделей, указанных выше. Из графиков видно, что чаще всего плотность распределения имеет один пик, вокруг которого сосредоточено основное количество результатов.

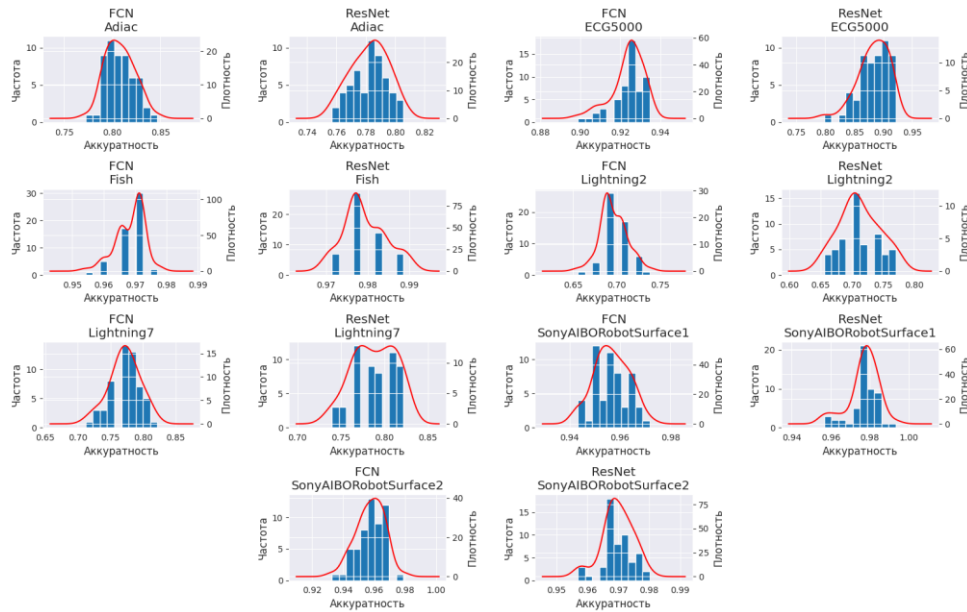


Рис. 1. Плотность распределения акkuratности классификации на различных датасетах

Проверка распределения акkuratности моделей на соответствие нормальному проводилась с целью выяснить, допустимо ли использование квантилей нормального закона для построения доверительных интервалов. Для оценки соответствия распределений были применены критерии Шапиро–Уилка (p -значение больше 0,05) и Андерсона–Дарлинга (уровень значимости 0,05). Результаты представлены в табл. 2. Модель FCN демонстрирует нормальность распределения на датасетах Adiac, SonyAIBORobotSurface1 и SonyAIBORobotSurface2 по обоим критериям. Для Lightning7 критерий Шапиро–Уилка дал положительный результат $p = 0,0813$, однако по критерию Андерсона–Дарлинга значение статистики 0,9266 превысило критический уровень 0,739, что указывает на отсутствие нормальности. На остальных наборах данных (ECG5000, Fish, Lightning2) оба критерия опровергли гипотезу нормальности.

Результаты модели ResNet показали меньшее соответствие нормальному распределению по критериям Шапиро–Уилка и Андерсона–Дарлинга, подтвердив его только на Adiac и SonyAIBORobotSurface2. По критерию Андерсона–Дарлинга распределение соответствовало нормальному для ECG5000, несмотря на низкое p -значение 0,0172 по Шапиро–Уилку. На остальных датасетах (Fish, Lightning2, Lightning7, SonyAIBORobotSurface1) оба критерия показали отклонение от нормального распределения.

Таким образом, распределения метрик моделей FCN и ResNet в большинстве случаев не являются нормальными. Обе модели на обоих критериях отклонили гипотезу нормальности на датасетах Fish и Lightning2. На датасетах Adiac и SonyAIBORobotSurface2 обе модели удовлетворяют условиям нормальности обоих тестов. В остальных комбинациях датасетов и моделей наблюдается расхождение.

Таблица 2

Нормальность распределения значений аккуратности, обученных на разных сидах

Сеть	Датасет	P-значение критерия Шапиро–Уилка	Критерий Андерсона–Дарлинга
FCN	Adiac	0,6573	0,3593
	ECG5000	0,0002	1,7574
	Fish	0,0000	5,1043
	Lightning2	0,0001	3,2711
	Lightning7	0,0813	0,9266
	SonyAIBORobotSurface1	0,3206	0,4098
	SonyAIBORobotSurface2	0,4734	0,3895
ResNet	Adiac	0,2832	0,4118
	ECG5000	0,0172	0,6345
	Fish	0,0000	3,6117
	Lightning2	0,0187	0,9820
	Lightning7	0,0025	1,3357
	SonyAIBORobotSurface1	0,0002	2,1110
	SonyAIBORobotSurface2	0,0913	0,5973

Новая метрика воспроизводимости. Наиболее часто используемые метрики для оценки качества моделей: аккуратность (accuracy), точность (precision), полнота (recall), F1-мера. В связи с наблюдаемым варьированием результатов при повторных запусках с использованием различных состояний генератора случайных чисел предлагается ввести некоторый способ учета при сравнении результатов.

Можно использовать среднее значение и стандартное отклонение, однако в данном случае отсутствует универсальность при варьирующемся количестве запусков на разных сидах. Поэтому имеет смысл ввести мета-метрику воспроизводимости:

$$RM(\lambda, n) = \mu_m - \lambda \frac{\sigma_m}{\sqrt{n}}$$

где $\lambda \geq 0$ – коэффициент штрафа за нестабильность, который нужно выбирать эмпирически, n – количество повторных запусков на различных инициализирующих сидах. Чем выше значение λ , тем сильнее учитывается влияние разброса результатов между запусками.

Предлагаемая формула структурно аналогична нижней границе доверительного интервала для среднего значения. Однако оригинальную формулу доверительного интервала нельзя использовать из-за того, что распределение значений метрик при обучении на различных инициализирующих сидах не является нормальным.

На основе предложенной мета-метрики RM и наиболее часто используемых в задаче классификации меток можно рассмотреть 4 производных метрики: (i) $RM_{acc}(\lambda, n)$ – метрика воспроизводимости аккуратности; (ii) $RM_{precision}(\lambda, n)$ – метрика воспроизводи-

сти точности; (iii) $RM_{\text{recall}}(\lambda, n)$ – метрика воспроизводимости полноты; (iv) $RM_{\text{F1}}(\lambda, n)$ – метрика воспроизводимости F1-меры. Аналогичным способом можно ввести метрику воспроизводимости для других метрик, представленных числовыми значениями.

Формализация выбора коэффициента

Для подбора коэффициента λ , определяющего величину штрафа в метрике воспроизводимости, рассматривается множество значений метрики качества, полученное из нескольких независимых запусков модели: $M = \{m_1, \dots, m_N\}$. Формируется набор X подмножеств x фиксированного размера n , для которого вычисляется значение $RM_x(\lambda, n)$. После вычисляется относительная ошибка

$$\delta_x = \frac{|RM(\lambda, n) - \min(x)|}{\min(x)}.$$

Коэффициент λ для фиксированного размера n определяется минимизацией средней ошибки по множеству подвыборок X :

$$\bar{\Delta}_n(\lambda) = \frac{1}{|X|} \sum_{x \in X} \delta_x(\lambda)$$

$$\lambda_n^* = \operatorname{argmin}_{\lambda \in A} \bar{\Delta}_n(\lambda),$$

где A – заданный диапазон допустимых значений λ , определенных в следующем разделе, а x – множество выбранных подвыборок размера n . Таким образом, λ_n^* – это такое значение коэффициента, при котором метрика RM в среднем наиболее близка к минимальному наблюдаемому значению метрики качества и, следовательно, наиболее строго и надёжно штрафует модели с высоким разбросом результатов.

Выбор параметров

Проведен эксперимент по выбору значения коэффициента λ . Для каждой пары «датасет–сеть» из первичной выборки, состоящей из 55 независимых запусков, сформировано множество значений воспроизводимости X . Далее для каждого фиксированного размера подвыборки $n \in \{5, 10, 15\}$ методом бутстрэппинга было сгенерировано по 1000 случайных подвыборок из X . Также были рассмотрены значения λ в диапазоне от 1 до 15 с шагом 0,1. Для каждой подвыборки и каждого λ вычислялась метрика воспроизводимости RM . Для оценки качества каждого λ были вычислены абсолютные значения относительной ошибки которая показывает, насколько близко оценка RM приближается к эмпирическому минимуму выборки X . Средняя величина Δ_n по 1000 значениям δ_n подвыборок и каждой пары «датасет–сеть» позволила определить λ , минимизирующее среднюю относительную ошибку при заданном n .

В табл. 3 представлены полученные значения $\lambda_5, \lambda_{10}, \lambda_{15}$ с соответствующими средними ошибками $\Delta_5, \Delta_{10}, \Delta_{15}$. Из таблицы виден рост значений и снижение значений при увеличении размера подвыборок. Также следует отметить, что для всех пар «датасет–сеть» соблюдается $\lambda_5 \leq \lambda_{10} \leq \lambda_{15}$, при этом $\Delta_5 \geq \Delta_{10} \geq \Delta_{15}$. При меньших n дисперсия оценок метрики возрастает, что соответствует условиям с наибольшей статистической неопределенностью. Это позволяет определить такое значение λ , при котором метрика воспроизводимости будет близка к эмпирическому минимуму даже при самом ограниченном числе запусков.

Таблица 3

Значение, λ при котором достигается минимальная разность с наименьшим значением

Датасет	Сеть	λ_5	Δ_5	λ_{10}	Δ_{10}	λ_{15}	Δ_{15}
Adiac	FCN	4,9	0,0125	7,4	0,0081	9,0	0,0060
Adiac	ResNet	4,1	0,0102	6,1	0,0069	7,6	0,0048
ECG5000	FCN	5,9	0,0118	9,1	0,0075	11,4	0,0060

Окончание табл. 3

Датасет	Сеть	λ_5	Δ_5	λ_{10}	Δ_{10}	λ_{15}	Δ_{15}
ECG5000	ResNet	7,0	0,0381	10,4	0,0268	12,3	0,0215
Fish	FCN	5,9	0,0056	8,8	0,0039	11,0	0,0029
Fish	ResNet	3,1	0,0023	4,8	0,0014	5,9	0,0011
Lightning2	FCN	6,0	0,0186	8,4	0,0125	9,9	0,0098
Lightning2	ResNet	3,6	0,0262	5,4	0,0167	6,7	0,0124
Lightning7	FCN	5,0	0,0281	7,8	0,0182	9,5	0,0135
Lightning7	ResNet	4,3	0,0199	6,1	0,0122	7,5	0,0094
SonyAIBORobotSurface1	FCN	3,8	0,0045	5,9	0,0027	7,3	0,0021
SonyAIBORobotSurface1	ResNet	4,5	0,0097	7,4	0,0059	9,5	0,0046
SonyAIBORobotSurface2	FCN	5,5	0,0083	8,1	0,0057	10,0	0,0043
SonyAIBORobotSurface2	ResNet	5,4	0,0051	7,3	0,0032	9,2	0,0025

Рекомендуемое значение λ вычисляется как средневзвешенное усредненных значений по размеру подвыборки для всех пар «датасет–сеть». Весовые коэффициенты $w_i = 1/\Delta$. Полученное значение составляет $\lambda=4,51$.

Результат

В табл. 4 представлены характеристики относительной ошибки, вычисленные для подвыборок размера 5 с использованием $\lambda=4,51$. Средние значения относительной ошибки находятся в диапазоне $\{-0,0071, 0,0460\}$, демонстрируя, что оценка RM в большинстве случаев близка к минимальному значению, полученному на полном множестве. Исключение составляет пара «ECG5000–ResNet», для которой зафиксировано максимальное среднее значение 0,046 и наибольшая дисперсия (0,0326), что указывает на нестабильность модели на данном датасете. На парах «Fish–ResNet» и «Lightning2–ResNet» средние ошибки являются отрицательными, что означает, что значение RM с предложенным λ в среднем превосходит эмпирический минимум на полном множестве. Медианные значения также остаются вблизи нуля, подтверждая отсутствие систематического смещения оценок в сторону переоценки или недооценки. Стандартное отклонение не превышает 0,0353. Таким образом, можно SonyAIBORobotSurface1 заключить, что предложенное значение параметра сглаживания обеспечивает устойчивое и достаточно точное приближение к нижней границе метрики воспроизводимости.

Таблица 4

Характеристики относительной ошибки RM и минимального значения

Датасет	Сеть	mean	std	min	median	max
Adiac	FCN	0,0082	0,0138	-0,0382	0,0093	0,0506
Adiac	ResNet	0,0017	0,0128	-0,0363	0,0007	0,0426
ECG5000	FCN	0,0122	0,0099	-0,013	0,0125	0,0299
ECG5000	ResNet	0,046	0,0326	-0,0493	0,0512	0,1168

Окончание табл. 4

Датасет	Сеть	mean	std	min	median	max
Fish	FCN	0,0062	0,0051	-0,009	0,0066	0,0179
Fish	ResNet	-0,0018	0,0035	-0,0125	-0,0024	0,0089
Lightning2	FCN	0,0192	0,0179	-0,0443	0,0199	0,075
Lightning2	ResNet	-0,0071	0,0353	-0,1095	-0,0073	0,1129
Lightning7	FCN	0,0205	0,0286	-0,0602	0,023	0,0962
Lightning7	ResNet	0,0031	0,0247	-0,0582	0,0032	0,0926
SonyAIBORobotSurface1	FCN	0,0003	0,0053	-0,0151	0,0008	0,0144
SonyAIBORobotSurface1	ResNet	0,0068	0,0091	-0,0196	0,0083	0,0229
SonyAIBORobotSurface2	FCN	0,0082	0,0083	-0,0169	0,0076	0,0292
SonyAIBORobotSurface2	ResNet	0,0038	0,0048	-0,0103	0,005	0,0143

В табл. 5 представлены наименьшее и среднее значение метрик, а также $RM(4.51, 55)$ на полном множестве для пар «датасет–сеть». Во всех парах значение RM , занимает промежуточное положение между минимальным и средним значением, что свидетельствует о сбалансированности оценки. Разность между RM и минимальным значением варьируется от 0,0049 (Fish–ResNet) до 0,0698 (ECG5000–ResNet). При этом разность между RM и средним значением варьируется от 0,0027 (Fish–FCN) до 0,0195 (Lightning2–ResNet1D), что значительно меньше, чем до минимального. При этом средняя разность RM и минимального значения составляет 0.0252, а среднего значения и RM – 0.0084, что говорит о том, что RM значительно ближе к среднему значению, чем к минимальному. Таким образом, предложенное значение $\lambda=4,51$ позволяет при малых n проводить оценку нижней границы значения метрики. При этом, при увеличении количества повторений, позволяет сохранять значение метрики, близкое к среднему.

Таблица 5

Сравнение наименьшего значения выборки, метрики RM и среднего значения

Датасет	Сеть	min	RM	mean
Adiac	FCN	0,7724	0,7987	0,8078
Adiac	ResNet	0,757	0,775	0,7826
ECG5000	FCN	0,8989	0,9192	0,924
ECG5000	ResNet	0,7982	0,868	0,8838
Fish	FCN	0,9543	0,9658	0,9685
Fish	ResNet	0,9714	0,9763	0,9793
Lightning2	FCN	0,6557	0,6871	0,6963
Lightning2	ResNet	0,6557	0,6929	0,7124
Lightning7	FCN	0,7123	0,7576	0,7718

Окончание табл. 5

Датасет	Сеть	min	RM	mean
Lightning7	ResNet	0,7397	0,7736	0,7883
SonyAIBORobotSurface1	FCN	0,9434	0,9525	0,9566
SonyAIBORobotSurface1	ResNet	0,9567	0,9723	0,977
SonyAIBORobotSurface2	FCN	0,9328	0,952	0,9575
SonyAIBORobotSurface2	ResNet	0,957	0,9669	0,9698

Таким образом, результирующая новая предлагаемая мета-метрика воспроизводимости:

$$RM(\lambda) = \mu_m - 4,51 \frac{\sigma_m}{\sqrt{n}}.$$

Валидация на DenseNet

Эксперименты по воспроизводимости дополнительно были проведены на тех же 7 датасетах с использованием сети DenseNet [20]. В таблице 6 представлены значения аккуратности и метрики RM для сети DenseNet с наибольшим и наименьшим стандартным отклонением аккуратности. Из таблицы видно, что при относительно небольшом стандартном отклонении 0,011 (датасет Fish) штраф также небольшой и среднее значение RM по 1000 подвыборок (размер 5) близко к среднему значению аккуратности. С другой стороны, на данных Lightning7 наблюдается большая нестабильность 0,033 и размер штрафа составил 0,064.

Таблица 6

Значения аккуратности и RM сети DenseNet

Датасет	Средняя аккуратность	Стандартное отклонение аккуратности	Размах	Среднее RM	Штраф
Fish	0,969	0,011	0,046	0,949	0,021
Lightning7	0,701	0,033	0,137	0,637	0,064

Заключение. Данное исследование подчеркивает значимость выбора случайного седа при обучении сверточных нейронных сетей, решающих задачи классификации временных рядов. Полученные результаты показывают, что варьирование состояния генератора случайных чисел может вызывать разницу в метрике качества – до 12% в отдельных случаях. Сравнение моделей, результаты которых получены на единственном запуске, могут являться некорректными. Демонстрируется необходимость систематической оценки устойчивости результатов.

В работе демонстрируется, что распределение значений аккуратности (ассигасу), полученных при множественных запусках модели, в большинстве случаев не подчиняется нормальному закону. Это делает традиционные статистические методы оценки (такие как доверительные интервалы) ненадежными. В связи с этим предложена новая мета-метрика воспроизводимости (RM), сочетающая среднее значение метрики и штраф за дисперсию, задаваемый параметром λ . Подбор λ выполнен эмпирически на основе бутстрэппинга, что позволило обеспечить устойчивость оценки даже при малом числе запусков. Анализ показал, что RM находится между минимальным и средним значением метрики при большом количестве повторений и ближе к минимальному при малом количестве запусков.

Таким образом, работа не только дополнительно подтверждает существование проблемы нестабильности моделей при различной инициализации, но и предлагает практически применимый способ её учета. Предложенная методика RM может стать шагом к стандартизации оценки воспроизводимости в машинном обучении, особенно в контексте сравнений между новыми архитектурами и методами.

В дальнейшем перспективным представляется исследование поведения RM-метрики в задачах регрессии, генерации и сегментации, а также её адаптация к случаям мультиклассовой и многозадачной классификации. Ещё одним направлением может стать автоматическая настройка λ на основе свойств данных или архитектуры модели.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Glorot X., Bengio Y. Understanding the difficulty of training deep feedforward neural networks, *Proceedings of the thirteenth international conference on artificial intelligence and statistics: JMLR Workshop and Conference Proceedings*, 2010, pp. 249-256.
2. Fellicious C., Weissgerber T., Granitzer M. Effects of random seeds on the accuracy of convolutional neural networks, *International Conference on Machine Learning, Optimization, and Data Science*. Cham: Springer International Publishing, 2020, pp. 93-102.
3. Krizhevsky A. et al. Learning multiple layers of features from tiny images, 2009.
4. Hu M. Y. et al. Latent state models of training dynamics, *arXiv preprint arXiv:2308.09543*, 2023.
5. Picard D. Torch. manual_seed (3407) is all you need: On the influence of random seeds in deep learning architectures for computer vision, *arXiv preprint arXiv:2109.08203*, 2021.
6. Krizhevsky A. et al. Learning multiple layers of features from tiny images, 2009.
7. Deng J. et al. Imagenet: A large-scale hierarchical image database, *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248-255.
8. Touvron H. et al. Augmenting convolutional networks with attention-based aggregation, *arXiv preprint arXiv:2112.13692*, 2021.
9. Guo Z. et al. A grey-box attack against latent diffusion model-based image editing by posterior collapse, *arXiv preprint arXiv:2408.10901*, 2024.
10. Picard D. Torch. manual_seed (3407) is all you need: On the influence of random seeds in deep learning architectures for computer vision, *arXiv preprint arXiv:2109.08203*, 2021.
11. Singhal V. et al. BANKSY unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis, *Nature genetics*, 2024, Vol. 56. No. 3, pp. 431-441.
12. Chen S. et al. Diffusiondet: Diffusion model for object detection, *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 19830-19843.
13. Wightman R., Touvron H., Jégou H. Resnet strikes back: An improved training procedure in timm // *arXiv preprint arXiv:2110.00476*. – 2021.
14. Åkesson J., Töger J., Heiberg E. Random effects during training: Implications for deep learning-based medical image segmentation, *Computers in Biology and Medicine*, 2024, Vol. 180, pp. 108944.
15. Karthik S. et al. If at first you don't succeed, try, try again: Faithful diffusion-based text-to-image generation by selection, *arXiv preprint arXiv:2305.13308*, 2023.
16. Touvron H. et al. Augmenting convolutional networks with attention-based aggregation, *arXiv preprint arXiv:2112.13692*, 2021.
17. Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.
18. Wang Z., Yan W., Oates T. Time series classification from scratch with deep neural networks: A strong baseline, *2017 International joint conference on neural networks (IJCNN)*. IEEE, 2017, pp. 1578-1585.
19. He K. et al. Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
20. Huang G. et al. Densely connected convolutional networks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700-4708.

Доброхвалов Максим Олегович – Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина); e-mail: night1337bot@gmail.com; г. Санкт-Петербург, Россия; тел.: +78122342682; кафедра математического обеспечения и применения ЭВМ; ORCID: 0000-0002-0571-5836.

Филатов Антон Юрьевич – Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина); e-mail: aifilatrov@etu.ru; г. Санкт-Петербург, Россия; тел.: +78122342682; к.т.н.; доцент кафедры математического обеспечения и применения ЭВМ; ORCID: 0000-0003-4298-8523.

Чегодаева Елизавета Александровна – Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина); e-mail: elizaveta.cheg@gmail.com; г. Санкт-Петербург, Россия; тел.: +78122342682; магистрант кафедры математического обеспечения и применения ЭВМ; ORCID: 0009-0005-9968-3108.

Dobrokhvalov Maksim Olegovich – Saint Petersburg Electrotechnical University "LETI"; e-mail: night1337bot@gmail.com; Saint Petersburg, Russia; phone: +78122342682; the Department of Software Engineering and Computer Applications; ORCID: 0000-0002-0571-5836.

Filatov Anton Yuryevich – Saint Petersburg Electrotechnical University "LETI"; e-mail: aifilatov@etu.ru; Saint Petersburg, Russia; phone: +78122342682; cand. of eng. sc.; associate professor Department of Software Engineering and Computer Applications; ORCID: 0000-0003-4298-8523.

Chegodaeva Elizaveta Aleksandrovna – Saint Petersburg Electrotechnical University "LETI"; e-mail: elizaveta.cheg@gmail.com; Saint Petersburg, Russia; phone: +78122342682; master's student of the Department of Software Engineering and Application of Computers; ORCID: 0009-0005-9968-3108.

Раздел II. Анализ данных, моделирование и управление

УДК 621.376

DOI 10.18522/2311-3103-2026-1-65-77

К.Е. Румянцев, Д.А. Цыщорин

МОДЕЛЬ ФОРМИРОВАНИЯ КЛЮЧЕВОЙ ПОСЛЕДОВАТЕЛЬНОСТИ В КОГЕРЕНТНОЙ ДВУХПОЛОСНОЙ ОПТИЧЕСКОЙ СИСТЕМЕ С МОДУЛЯЦИЕЙ ФАЗОМАНИПУЛИРОВАННЫМИ РАДИОСИГНАЛАМИ

Когерентная оптическая связь с использованием квадратурной фазовой манипуляции (QPSK) обеспечивает высокую скорость передачи данных синфазными и квадратурными (IQ) форматами модуляции на большие расстояния. Квадратурный оптический модулятор с двумя интерферометрами Маха-Цендера в двухтактной конфигурации при одновременном фазовом сдвиге в его плечах позволяет существенно улучшить технические характеристики и обеспечить электромагнитную совместимость трактов передачи цифровых сигналов. Двухтактная конфигурация оптического модулятора является основой для организации высокоскоростной и помехозащищённой передачи данных. Результаты исследований доказывают, что после определённой коррекции известных структур когерентной оптической передачи система может быть использована для распределения ключа. Представлена структура передающей станции когерентной двухполосной оптической системы формирования ключевой последовательности. Модель формирования оптического излучения с модуляцией фазоманипулированным радиосигналом осуществляется на основе квадратурного оптического модулятора на двух параллельно включённых интерферометрах Маха-Цендера с постоянными напряжениями смещения на всех плечах управляющих электродов. Формирование ключевой последовательности реализуется посредством модуляции синфазной и квадратурной составляющих радиосигналов на поднесущей частоте. Набор двух коэффициентов умножения амплитуд для каждого бита задаётся электронным устройством кодирования. Коэффициенты представляют сигналы, модулирующие по амплитуде соответственно синфазную и квадратурную составляющие поднесущей частоты. Формируемое оптическое излучение на выходе модулятора обеспечивает распределение ключа (задание значений нулевого «0» и единичного «1» битов в двоичной системе счисления) по протоколу с четырьмя состояниями фазы в прямоугольном и диагональном базисах. Предложен модуль управления кодированием синфазной и квадратурной составляющих радиосигнала на поднесущей частоте. Применение квадратурной фазовой манипуляции (QPSK) обеспечивает помехоустойчивость к внешним воздействиям при высокой скорости передачи данных.

Распределение ключа; когерентное излучение; конфигурация двух интерферометров Маха-Цендера; формирование двухполосного оптического излучения; квадратурный фазоманипулированный радиосигнал; поднесущая частота.

K.Y. Rumyantsev, D.A. Tsytsorin

MODEL OF KEY SEQUENCE GENERATION IN A COHERENT DUAL-BAND OPTICAL SYSTEM WITH PHASE-SHIFT KEYED RADIO SIGNAL MODULATION

Coherent optical communication using quadrature phase-shift keying (QPSK) provides high-speed data transmission using in-phase and quadrature (IQ) modulation formats over long distances. A quadrature optical modulator with two Mach-Zehnder interferometers in a two-way configuration with simultaneous phase shift in its arms significantly improves the technical characteristics and ensures the electromagnetic compatibility of digital signal transmission paths. The two-cycle configuration of an optical modulator is the basis for high-speed and noise-resistant data transmission. The research results prove that, after certain

corrections to the known coherent optical transmission structures, the system can be used for key distribution. The structure of a coherent two-band optical key sequence formation system is presented. The model for generating optical radiation with modulation by a phase-shifted radio signal is based on a quadrature optical modulator with two parallel Mach-Zehnder interferometers with constant bias voltages on all control electrode arms. Key sequence formation is implemented by modulating the in-phase and quadrature components of radio signals at the subcarrier frequency. A set of two amplitude multiplication coefficients for each bit is specified by the electronic encoding device. The coefficients represent the signals that modulate the amplitude of the in-phase and quadrature components of the subcarrier frequency, respectively. The optical radiation generated at the modulator output provides key distribution (setting the values of zero "0" and one "1" bits in the binary number system) according to a protocol with four phase states in the rectangular and diagonal bases. A module for controlling the encoding of the in-phase and quadrature components of a radio signal at a subcarrier frequency is proposed. The use of quadrature phase shift keying (QPSK) provides noise immunity to external influences at high data transfer rates.

Key distribution, coherent radiation; configuration of two Mach-Zehnder interferometers; formation of dual-band optical radiation; quadrature phase-shift keyed radio signal; subcarrier frequency.

Введение. Технология когерентной оптической связи позволяет обеспечить передачу информации на большие расстояния с высокой пропускной способностью [1, 2]. В частности, применение квадратурной фазовой манипуляции (QPSK – quadrature phase-shift keying или 4-PSK) обеспечивает высокую скорость передачи данных простым, надёжным и экономичным способом [3]. Передача нескольких бит, как символа, реализуется простыми синфазными и квадратурными (IQ) форматами модуляции [4–8].

Для увеличения пропускной способности в когерентных оптических системах используется мультиплексирование [9–11], когда манипулированные разными потоками данных радиосигналы на нескольких поднесущих частотах модулируют излучение одного оптического квантового генератора (ОКГ). Одно из преимуществ мультиплексирования на поднесущих частотах (SCM – Subcarrier Multiplexing) состоит в том, что устройства СВЧ для формирования и обработки радиосигналов более стабильны, чем их оптические аналоги [12].

Перенос процесса модуляции во внешний оптический модулятор позволяет снизить влияние амплитудного шума лазерного источника и повысить скорость переключения. Применение интерферометра Маха-Цендера (ИМЦ) [13, 14] в качестве внешнего модулятора позволяет обеспечить фазовую модуляцию в высокоскоростных оптических системах передачи.

Интерферометр Маха-Цендера в интегральном исполнении включает оптический ответвитель Y-типа (делитель) и оптический сумматор Y-типа, соединённые прямолинейными участками волноводов (плечами ИМЦ) на монокристалле [15, 16]. Интерферометр изготавливается из титана, который диффундируется в слой подложки из монокристалла ниобата лития (LiNbO_3) Z-срезы. Кристалл обладает сильным электрооптическим эффектом, так что если к нему приложить напряжение, то показатель преломления будет изменён [17]. Изменение показателя преломления под действием эффекта Поккельса при приложении к монокристаллу напряжения приводит к изменению оптической длины и к фазовому сдвигу [18].

Использование интерферометра Маха-Цендера при одновременном фазовом сдвиге в его двух плечах позволяет существенно улучшить технические характеристики и обеспечить электромагнитную совместимость трактов передачи цифровых сигналов [19]. Дополнительные возможности даёт применение двух параллельно включённых интерферометров Маха-Цендера [20].

Квадратурный оптический модулятор с двумя интерферометрами Маха-Цендера в двухтактной конфигурации является основным элементом для организации высокоскоростной и помехозащищённой передачи данных [21].

В [22] проанализирована система когерентной оптической передачи с одной боковой полосой при модуляции оптического излучения фазоманипулированным радиосигналом с использованием оптического квадратурного модулятора и формированием импульсов с M-ричной радиочастотой. Генератор псевдослучайных битовых последова-

тельностью генерирует модулирующий сигнал со скоростью передачи до 10 Гбит/с. Битовая последовательность преобразуется в М-ричную последовательность символов из двоичных сигналов с использованием фазовой манипуляции (PSK). Генератор формирует многоуровневые импульсы в соответствии с последовательностью М-ричных входных символов. Электрические импульсы подаются на управляющие электроды электрооптического модулятора на основе интерферометра Маха–Цендера (ИМЦ). Напряжения смещения обеспечивают работу интерферометра в квадратурной точке. Лазером излучается непрерывная оптическая волна с длиной волны 1550 нм.

Представленные ниже результаты исследований доказывают, что после определённой коррекции описанной ранее структуры когерентной оптической передачи система может быть использована для распределения ключа.

Формирование оптического излучения с модуляцией фазоманипулированным радиосигналом на основе двух параллельно включённых интерферометров Маха–Цендера. Проанализируем структуру квадратурного оптического модулятора на двух параллельно включённых интерферометров Маха–Цендера ИМЦ1 и ИМЦ2, как показано на рис. 1. Здесь постоянные напряжения смещения подаются на кристаллы соответственно в верхних $U_{см11}$, $U_{см12}$ и нижних $U_{см21}$, $U_{см22}$ плечах интерферометров ИМЦ1 и ИМЦ2.

Пусть оптический квантовый генератор (ОКГ) генерирует непрерывное немодулированное монохроматическое излучение с мощностью $P_{опт}$, частотой $f_{опт}$, начальной фазой $\varphi_{опт}$ и мгновенной фазой $\theta_{опт} = 2\pi f_{опт}t + \varphi_{опт}$.

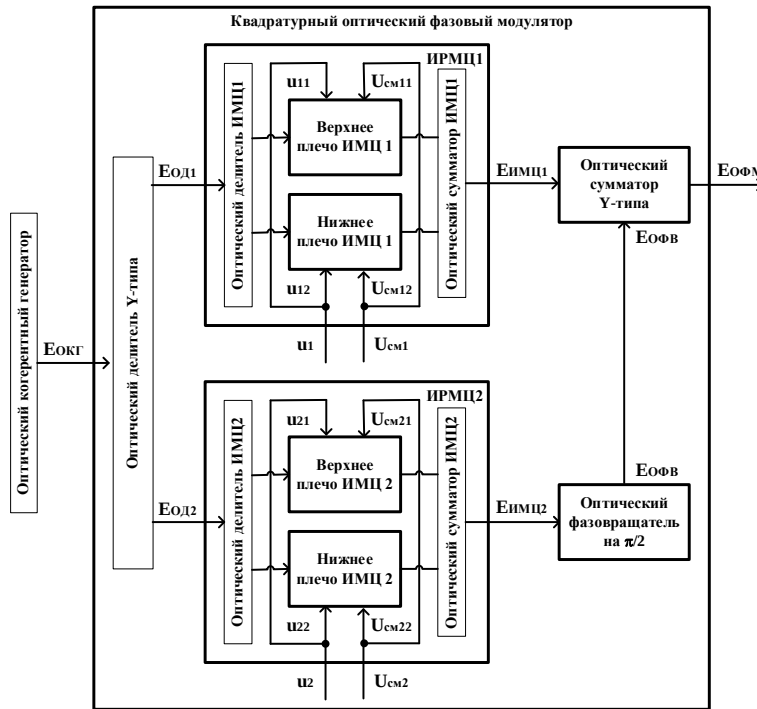


Рис. 1. Квадратурный оптический модулятор на двух интерферометрах

Мгновенное значение электромагнитного поля оптического колебания ОКГ с амплитудой $E_{опт} = \sqrt{2 \cdot P_{опт}}$ описывается формулой

$$E_{ОКГ}(t) = E_{опт} \cdot \cos(\theta_{опт}) = \sqrt{2 \cdot P_{опт}} \cdot \cos(\theta_{опт}). \quad (1)$$

В формуле учтено, что мгновенное значение мощности ОКГ (лазера, лазерного диода) определяется квадратом поля оптического излучения. Усреднение по периоду оптического колебания $P_{ОКГ}(t)$ даёт среднюю мощность оптического поля лазера $P_{опт}$.

Непрерывное монохроматическое излучение ОКГ (1) посредством оптического делителя (ОД, оптического разветвителя Y-типа) разделяется на две оптические волны $E_{ОД1}(t)$ и $E_{ОД2}(t)$.

$$E_{ОД1}(t) = E_{ОД2}(t) = E_{ОД}(t) = \frac{E_{ОКГ}(t)}{2} = \sqrt{\frac{P_{ОПТ}}{2}} \cdot \cos(\theta_{ОПТ}) \quad (2)$$

с равными амплитудами $\sqrt{P_{ОПТ}}/2$ на входах первого (ИМЦ1) и второго (ИМЦ2) интерферометров Маха–Цендера.

Отметим, что электромагнитные поля $E_{ОД1}(t)$ и $E_{ОД2}(t)$ с выходов оптического делителя дополнительно делятся поровну между нижними и верхними плечами (кристаллами) интерферометров Маха–Цендера благодаря оптическим делителям. Таким образом, амплитуды электромагнитных полей на входах в каждое плечо интерферометров $\sqrt{P_{ОПТ}}/8$ в четыре раза меньше амплитуды оптического колебания ОКГ.

Первый верхний интерферометр Маха–Цендера ИМЦ1 содержит в каждом плече пары отдельных верхних и нижних управляющих электродов, на которые подаются напряжения $u_{ИМЦ11}(t) = u_{11}(t) + U_{см11}$ и $u_{ИМЦ12}(t) = u_{12}(t) + U_{см12}$, которые управляют электрическими полями в плечах интерферометра. Выбор рабочей точки модулятора достигается посредством приложения к электродам напряжений смещения $U_{см11}$ и $U_{см12}$, которые обеспечивают заданный сдвиг фаз между оптическими волнами в плечах интерферометра.

На выходе первого интерферометра Маха–Цендера происходит интерференция сдвинутых по фазе оптических волн. Выходная оптическая волна верхнего первого интерферометра ИМЦ1 с коэффициентами передачи мощности в каждом плече $K_{р.ИМЦ1}$ описывается формулой

$$E_{ИМЦ1}(t) = \sqrt{K_{р.ИМЦ1}} \cdot \frac{E_{ОКГ}(t)}{4} \cdot \left\{ \exp \left[j\pi \frac{u_{11}(t) + U_{см11}}{U_{\pi}} \right] + \exp \left[j\pi \frac{u_{12}(t) + U_{см12}}{U_{\pi}} \right] \right\},$$

где U_{π} – полуволновое напряжение, необходимое для сдвига фазы оптического колебания на π .

Деление на 4 в формуле возникает из-за того, что амплитуда оптического излучения лазера делится поровну последовательно сначала в оптическом делителе, а затем между верхним и нижним плечами интерферометра.

Согласно представленной на рис. 1 схеме квадратурного оптического фазового модулятора на электроды кристаллов, определяющих плечи интерферометра, подаются напряжения одинаковой полярности, причём

$$u_{11}(t) = u_{12}(t) = u_1(t); \\ U_{см11} = U_{см12} = U_{см1}.$$

При этом на выходе верхнего интерферометра ИМЦ1 формируется оптическая волна

$$E_{ИМЦ1}(t) = \sqrt{K_{р.ИМЦ1}} \cdot \frac{E_{ОКГ}(t)}{2} \cdot \exp \left[j \frac{\pi}{U_{\pi}} u_1(t) + j \frac{\pi}{U_{\pi}} U_{см1} \right]. \quad (3)$$

На выходе первого интерферометра имеем созвездие сигналов бинарной фазовой манипуляции BPSK.

Аналогично, оптическая волна на выходе нижнего интерферометра Маха–Цендера ИМЦ2 с равными коэффициентами передачи мощности в каждом плече $K_{р.ИМЦ2}$ и при выполнении условий

$$u_{21}(t) = u_{22}(t) = u_2(t) \\ U_{см21} = U_{см22} = U_{см2}$$

описывается формулой

$$E_{ИМЦ2}(t) = \sqrt{K_{р.ИМЦ2}} \cdot \frac{E_{ОКГ}(t)}{2} \cdot \exp \left[j \frac{\pi}{U_{\pi}} u_2(t) + j \frac{\pi}{U_{\pi}} U_{см2} \right]. \quad (4)$$

Видно, что на выходе второго (нижнего) интерферометра также формируется созвездие сигналов бинарной фазовой манипуляции BPSK.

К выходу нижнего интерферометра ИМЦ2 подключается оптический фазовращатель (ОФВ) на $\pi/2$ с коэффициентом передачи мощности $K_{P.ОФВ}$. С учётом (4) на выходе оптического фазовращателя формируется оптическая волна

$$E_{ОФВ}(t) = j \cdot \sqrt{K_{P.ОФВ}K_{P.ИМЦ2}} \cdot \frac{E_{ОКГ}(t)}{2} \cdot \exp \left[j \frac{\pi}{U_{\pi}} u_2(t) + j \frac{\pi}{U_{\pi}} U_{см2} \right]. \quad (5)$$

Сравнение формул (4) и (5) показывает, что благодаря фазового сдвига на $\pi/2$ излучения, поступающие на входные порты оптического сумматора с коэффициентом передачи мощности $K_{P.ОС}$, будут повернуты относительно друг друга на $\pi/2$.

На выходе оптического сумматора (ОС), который является выходом квадратурного оптического фазового модулятора (ОФМ), оптическая волна описывается формулой

$$E_{ОФМ}(t) = \sqrt{K_{P.ОС}} \cdot [E_{ИМЦ1}(t) + E_{ОФВ}(t)].$$

На практике обеспечивается равенство коэффициентов передачи мощности в плечах интерферометров Маха-Цендера $K_{P.ИМЦ1} = K_{P.ИМЦ2} = K_{P.ИМЦ}$. Потери в оптических фазовращателях обладают очень низкими значениями (не превышают 1 дБ), что позволяет принять $K_{P.ОС} = 1$.

С учётом (3) и (5) при $K_{P.ОС} = 1$ и равенстве коэффициентов передачи мощности $K_{P.ИМЦ1} = K_{P.ИМЦ2} = K_{P.ИМЦ}$ находим

$$E_{ОФМ}(t) = \sqrt{K_{P.ОС}K_{P.ИМЦ}} \cdot \frac{E_{ОКГ}(t)}{2} \cdot \left\{ \exp \left[j \frac{\pi}{U_{\pi}} u_1(t) + j \frac{\pi}{U_{\pi}} U_{см1} \right] + j \cdot \exp \left[j \frac{\pi}{U_{\pi}} u_2(t) + j \frac{\pi}{U_{\pi}} U_{см2} \right] \right\}.$$

В качестве рабочей точки выбирается квадратурная точка передаточной функции мощности фазового модулятора на интерферометре Маха-Цендера, обеспечивающая контролируемый сдвиг фаз на $-\pi/2$ между оптическими волнами в плечах интерферометра посредством приложения к электродам напряжений смещения $U_{см1} = U_{см2} = -U_{\pi}/2$. Это обеспечивает работу фазового модулятора на линейном участке передаточной функции мощности, представляющей зависимость мощности излучения на выходе идеального, симметричного модулятора от разности фаз волн в плечах интерферометра [15]. Причём рост напряжения на кристалле интерферометра влечёт увеличение фазового сдвига оптического излучения.

При $U_{см1} = U_{см2} = -U_{\pi}/2$ колебания оптической волны на выходе квадратурного оптического фазового модулятора описываются формулой

$$E_{ОФМ}(t) = \sqrt{K_{P.ОС}K_{P.ИМЦ}} \cdot \frac{E_{вх}(t)}{2} \cdot \left\{ -j \cdot \exp \left[j \frac{\pi}{U_{\pi}} u_1(t) \right] + \exp \left[j \frac{\pi}{U_{\pi}} u_2(t) \right] \right\}. \quad (6)$$

Из формулы (6) видно, что генерируется сигнал с квадратурной фазовой манипуляцией QPSK.

Формирование ключевой последовательности посредством модуляции синфазной и квадратурной составляющих радиосигнала. Пусть передающая станция включает задающий электронный генератор гармонического колебания на поднесущей радиочастоте (РЧ) $f_{рч}$ с амплитудой $U_{рч}$, начальной фазой $\varphi_{рч}$ и мгновенной фазой $\theta_{рч} = 2\pi f_{рч}t + \varphi_{рч}$:

$$u_{рч}(t) = U_{рч} \cos(2\pi f_{рч}t + \varphi_{рч}) = U_{рч} \cos(\theta_{рч}). \quad (7)$$

Формирование на передающей станции j -го бита в ключевой последовательности во временном интервале $t \in [t_j; t_j + \tau_{бит}]$ предусматривает генерацию в момент $t = t_j$ двух чисел $x_1(t_j)$ и $x_2(t_j)$, значения которых могут случайным образом принимать только два значения «-1» или «+1». Величина $\tau_{бит}$ определяет длительность двоичного бита.

Предлагаемый протокол формирования ключевой последовательности использует радиосигналы в двух базисах – в прямоугольном «+» ($k=1$) и диагональном «×» ($k=2$). Во время формирования j -го бита радиосигналы на поднесущей радиочастоте $f_{рч}$ задаются набором коэффициентов умножения амплитуд

$$I_{kj} = x_1(t_j), t \in [t_j; t_j + \tau_{\text{бит}}], k = 1; 2;$$

$$Q_{kj} = x_2(t_j), t \in [t_j; t_j + \tau_{\text{бит}}], k = 1; 2$$

соответственно для синфазных и квадратурных составляющих радиосигнала.

Из колебания (7) формируются два радиосигнала с синфазными $u_{\text{рч},I}(t)$ и квадратурными $u_{\text{рч},Q}(t)$ колебаниями, причём:

$$u_{\text{рч},I}(t) = U_{\text{рч}} \cos(\theta_{\text{рч}}),$$

$$u_{\text{рч},Q}(t) = U_{\text{рч}} \sin(\theta_{\text{рч}}).$$

Набор коэффициентов умножения амплитуд I_{kj} и Q_{kj} для каждого j -го бита задаётся электронным устройством кодирования. Компоненты I_{kj} и Q_{kj} представляют сигналы, модулирующие по амплитуде соответственно синфазную $u_{\text{рч},I}(t) = U_{\text{рч}} \cos(\theta_{\text{рч}})$ и квадратурную $u_{\text{рч},Q}(t) = U_{\text{рч}} \sin(\theta_{\text{рч}})$ составляющие поднесущей частоты.

Поскольку последующий анализ проводится применительно к формированию одного j -го бита, то, для упрощения записи, полагается $I_{kj} = I_k$ и $Q_{kj} = Q_k$. Тогда после кодирования для математического описания двух радиосигналов во время формирования j -го бита в интервале $t \in [t_j; t_j + \tau_{\text{бит}}]$ справедливы формулы

$$\begin{aligned} u_{k1}(t) &= I_k \cdot U_{\text{рч}} \cdot \cos(\theta_{\text{рч}}) - Q_k \cdot U_{\text{рч}} \cdot \sin(\theta_{\text{рч}}); \\ u_{k2}(t) &= I_k \cdot U_{\text{рч}} \cdot \cos(\theta_{\text{рч}}) + Q_k \cdot U_{\text{рч}} \cdot \sin(\theta_{\text{рч}}). \end{aligned} \quad (8)$$

Значения нулевого «0» и единичного «1» битов в двоичной системе счисления задаются в прямоугольном базисе генерации (индекс «+») соответственно наборами значений коэффициентов умножения амплитуд радиосигналов

$$\begin{aligned} "0_+" &\rightarrow \{I_1; Q_1\} = \{+1; +1\}; \\ "1_+" &\rightarrow \{I_1; Q_1\} = \{-1; -1\}. \end{aligned} \quad (9)$$

Значения нулевого «0» и единичного «1» битов задаются в диагональном базисе генерации (индекс «×») другим набором значений коэффициентов умножения амплитуд радиосигналов

$$\begin{aligned} "0_\times" &\rightarrow \{I_2; Q_2\} = \{+1; -1\}; \\ "1_\times" &\rightarrow \{I_2; Q_2\} = \{-1; +1\}. \end{aligned} \quad (10)$$

Формулы (6), (9) и (10) показывают, что структура оптического квадратурного модулятора на двух параллельно включённых интерферометрах Маха-Цендера с постоянными напряжениями смещения на всех плечах управляющих электродах, представленная на рис. 1, обеспечивает формирование оптического излучения с модуляцией фазоманипулированным радиосигналом на поднесущей частоте. Причём формируемое оптическое излучение на выходе модулятора обеспечивает распределение ключа по протоколу с четырьмя состояниями фазы в прямоугольном и диагональном базисах (рис. 2).

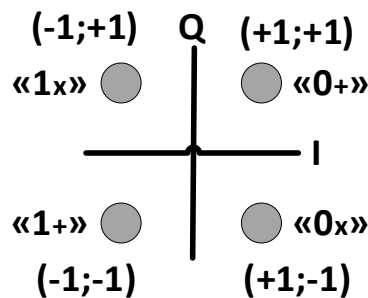


Рис. 2. Сигнальные созвездия фаз на выходе квадратурного модулятора

Кодирование синфазной и квадратурной составляющих. Остановимся на процессе формирования на передающей станции j -го бита в ключевой последовательности во временном интервале $t \in [t_j; t_j + \tau_{\text{бит}}]$. Для этого передающая станция должна включать модуль управления кодированием синфазной и квадратурной составляющих радиосигнала на поднесущей частоте. Возможный вариант модуля управления представлен на рис. 3.

Формирование j -го бита в ключевой последовательности в интервале $t \in [t_j; t_j + \tau_{\text{бит}}]$ предусматривает генерацию двумя генераторами случайных чисел (ГСЧ) в момент $t = t_j$ двух чисел $x_1(t_j)$ и $x_2(t_j)$, значения которых случайным образом принимают одно из двух значений «-1» или «+1».

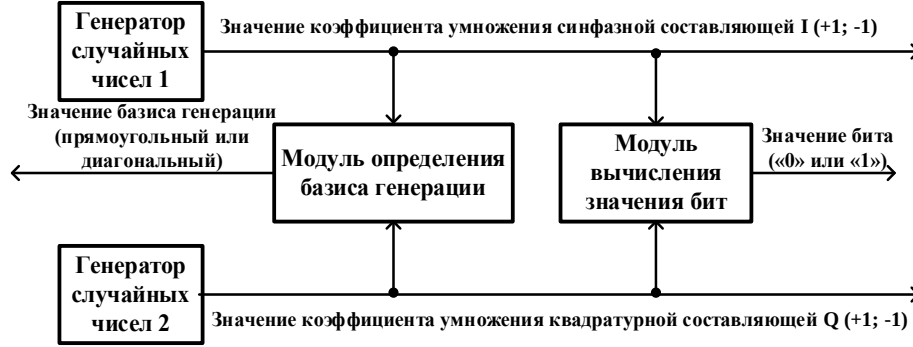


Рис. 3. Структурная схема электронного устройства кодирования синфазной и квадратурной составляющих радиосигнала на поднесущей частоте

Модуль определения базиса генерации на основании формул (9) и (10) вычисляет, какой базис генерации использован для кодирования состояния фазы радиосигнала. Эта информация необходима для получения просеянной ключевой последовательности из сырой последовательности.

Формулы (9) и (10) позволяют модулю вычисления значения бит определить значение (логического «0» или логической «1») бита генерируемого передающей станцией бита. Информация заносится в сырую ключевую последовательность передающей станции.

Структура передающей станции когерентной оптической системы распределения ключа символами. На рис. 4 приведена структурная схема передающей станции когерентной системы с амплитудной модуляцией оптического излучения радиосигналом. Для распределения ключа символами используется квадратурная фазовая манипуляция радиосигнала на поднесущей частоте.

При квадратурной фазовой манипуляции необходимы два колебания $u_{\text{рч},I}(t)$ и $u_{\text{рч},Q}(t)$ с одинаковой поднесущей радиочастотой $f_{\text{рч}}$, но сдвинутые по фазе на 90° .

Это достигается использованием электронного генератора синусоидальных сигналов поднесущей радиочастоты. На его выходе генерируется гармоническое колебание

$$u_{\text{рч},I} = U_{\text{рч}} \cos(\theta_{\text{рч}}) \quad (11)$$

с амплитудой $U_{\text{рч}}$, поднесущей частотой $f_{\text{рч}}$, начальной фазой $\varphi_{\text{рч}}$ и мгновенной фазой $\theta_{\text{рч}} = 2\pi f_{\text{рч}} t + \varphi_{\text{рч}}$.

На выходе электронного фазовращателя (ЭФВ) с коэффициентом передачи $K_{\text{ЭФВ}} = 1$ формируется гармоническое колебание

$$u_{\text{рч},Q}(t) = U_{\text{рч}} \sin(\theta_{\text{рч}}). \quad (12)$$

На выходах первого и второго электронных устройств перемножения (ЭУП) с равными коэффициентами передачи напряжения $K_{\text{ЭУП}}$ будут действовать радиосигналы с одинаковыми амплитудами $K_{\text{ЭУП}} U_{\text{рч}}$:

$$\begin{cases} u_{\text{ЭУП1}}(t) = K_{\text{ЭУП}} I_k u_{\text{рч},I} = K_{\text{ЭУП}} U_{\text{рч}} I_k \cos(\theta_{\text{рч}}); \\ u_{\text{ЭУП2}}(t) = K_{\text{ЭУП}} Q_k u_{\text{рч},Q} = K_{\text{ЭУП}} U_{\text{рч}} Q_k \sin(\theta_{\text{рч}}). \end{cases} \quad (13)$$

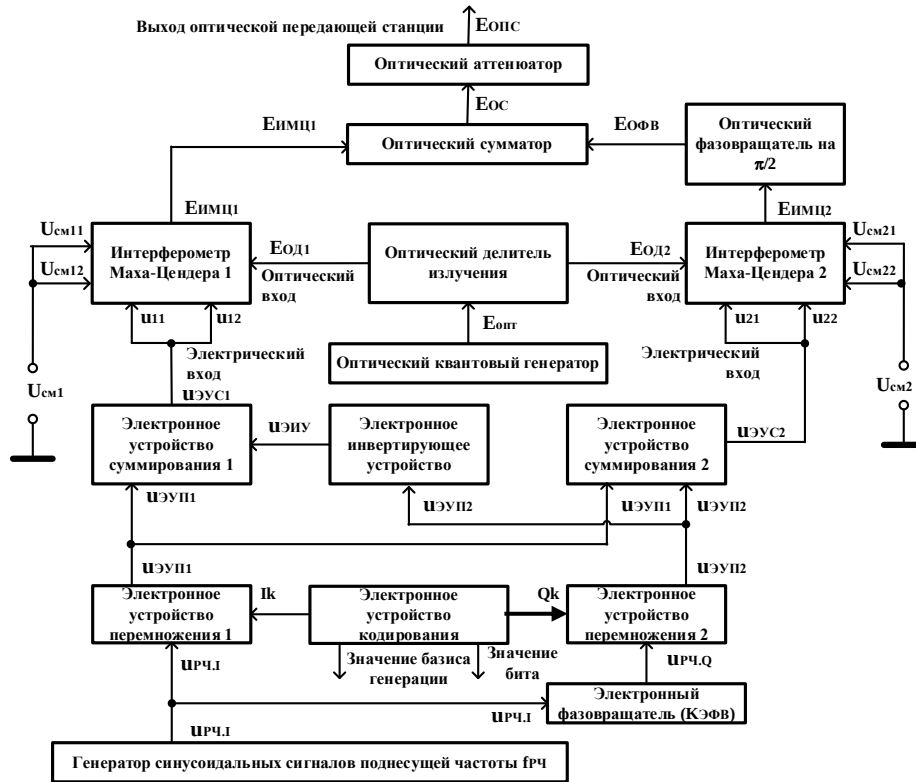


Рис. 4. Структурная схема передающей станции когерентной системы распределения ключа символами с амплитудной модуляцией оптического излучения радиосигналом

Отметим, что случайные значения коэффициентов I_k и Q_k , генерируемых электронным устройством кодирования, однозначно определяют базис генерации и значение бита согласно условий (9) и (10).

На выходе электронного инвертирующего устройства (ЭИУ) с коэффициентом передачи напряжения $K_{ЭИУ}$ действует радиосигнал $u_{ЭИУ}(t) = -u_{ЭУП2}(t) = -K_{ЭУП}Q_k u_{pч,Q}$ с амплитудой $K_{ЭИУ}K_{ЭУП}U_{pч}$:

$$u_{ЭИУ}(t) = -K_{ЭИУ}K_{ЭУП}U_{pч}Q_k \sin(\theta_{pч}). \quad (14)$$

На выходе первого аналогового электронного устройства суммирования (ЭУС1) с коэффициентом передачи напряжения $K_{ЭУС1}$ формируется радиосигнал

$$u_{ЭУС1}(t) = K_{ЭУС1}[u_{ЭУП1}(t) + u_{ЭИУ}(t)].$$

Поскольку электронное инвертирующее устройство легко может быть реализовано с коэффициентом передачи напряжения $K_{ЭИУ} \cong 1$, то

$$u_{ЭУС1}(t) = K_{ЭУС1}K_{ЭУП}U_{pч}[I_k \cos(\theta_{pч}) - Q_k \sin(\theta_{pч})]. \quad (15)$$

На выходе второго аналогового электронного устройства суммирования (ЭУС2) с коэффициентом передачи напряжения $K_{ЭУС2}$ формируется радиосигнал

$$u_{ЭУС2}(t) = K_{ЭУС2}[u_{ЭУП1}(t) + u_{ЭУП2}(t)].$$

С учётом (13) находим

$$u_{ЭУС2}(t) = K_{ЭУС2}K_{ЭУП}U_{pч}[I_k \cos(\theta_{pч}) + Q_k \sin(\theta_{pч})]. \quad (16)$$

Сигналы с выходов аналоговых электронных устройств суммирования (ЭУС1) $u_{ЭУС1}(t)$ и (ЭУС2) $u_{ЭУС2}(t)$ с равными коэффициентами передачи напряжения $K_{ЭУС1} = K_{ЭУС2} = K_{ЭУС}$ подаются на электрические входы соответственно первого и второ-

го интерферометров Маха-Цендера с равными коэффициентами передачи мощности в каждом плече интерферометра $K_{P,ИМЦ1} = K_{P,ИМЦ2} = K_{P,ИМЦ}$. На оптические входы интерферометров подаётся непрерывное лазерное излучение с постоянной мощностью $P_{опт}(t) = P_{опт}$. Напряжённость электрического поля лазерного источника определяется формулой (1).

Оптический делитель излучения с равными коэффициентами деления обеспечивает подачу в каждое плечо интерферометра Маха-Цендера излучений с одинаковой напряжённостью (2).

Первый интерферометр Маха-Цендера имеет отдельные верхние и нижние электроды, на которые подаются напряжения $u_{ИМЦ11}(t)$ и $u_{ИМЦ12}(t)$, которые управляют электрическими полями в плечах интерферометра, причём

$$\begin{aligned} u_{ИМЦ11}(t) &= u_{11}(t) + U_{см11} = u_{ЭУС1}(t) - U_{\pi}/2, \\ u_{ИМЦ12}(t) &= u_{12}(t) + U_{см12} = u_{ЭУС1}(t) - U_{\pi}/2. \end{aligned}$$

С учётом (15)

$$u_{ИМЦ11}(t) = u_{ИМЦ12}(t) = K_{ЭУС}K_{ЭУП}U_{рч}[I_k \cos(\theta_{рч}) - Q_k \sin(\theta_{рч})] - U_{\pi}/2.$$

Выходная оптическая волна первого интерферометра Маха-Цендера с коэффициентом передачи мощности $K_{P,ИМЦ}$ описывается формулой

$$E_{ИМЦ1}(t) = -j \cdot \sqrt{K_{P,ИМЦ}} \cdot \frac{E_{ОКГ}(t)}{2} \cdot \rightarrow \leftarrow \exp \left[j\pi K_{ЭУС1}K_{ЭУП} \frac{U_{рч}}{U_{\pi}} [I_k \cos(\theta_{рч}) - Q_k \sin(\theta_{рч})] \right]. \quad (17)$$

Второй интерферометр Маха-Цендера имеет отдельные верхние и нижние электроды, на которые подаются напряжения $u_{ИМЦ21}(t)$ и $u_{ИМЦ22}(t)$, причём

$$\begin{aligned} u_{ИМЦ21}(t) &= u_{21}(t) + U_{см21} = u_{ЭУС2}(t) - U_{\pi}/2, \\ u_{ИМЦ22}(t) &= u_{22}(t) + U_{см22} = u_{ЭУС2}(t) - U_{\pi}/2. \end{aligned}$$

С учётом (16) находим

$$u_{ИМЦ21}(t) = u_{ИМЦ22}(t) = K_{ЭУС}K_{ЭУП}U_{рч}[I_k \cos(\theta_{рч}) + Q_k \sin(\theta_{рч})] - U_{\pi}/2.$$

Выходная оптическая волна на выходе второго интерферометра Маха-Цендера с коэффициентом передачи мощности $K_{P,ИМЦ}$ описывается формулой

$$E_{ИМЦ2}(t) = -j \cdot \sqrt{K_{P,ИМЦ}} \cdot \frac{E_{ОКГ}(t)}{2} \cdot \rightarrow \leftarrow \exp \left[j\pi K_{ЭУС}K_{ЭУП} \frac{U_{рч}}{U_{\pi}} [I_k \cos(\theta_{рч}) + Q_k \sin(\theta_{рч})] \right]. \quad (18)$$

После оптического фазовращателя (ОФВ) на $\pi/2$ с коэффициентом передачи мощности $K_{P,ОФВ}$ будет сформировано излучение с напряжённостью

$$E_{ОФВ}(t) = \sqrt{K_{P,ОФВ}K_{P,ИМЦ}} \cdot \frac{E_{ОКГ}(t)}{2} \cdot \rightarrow \leftarrow \exp \left[j\pi K_{рч} \frac{U_{рч}}{U_{\pi}} [I_k \cos(\theta_{рч}) + Q_k \sin(\theta_{рч})] \right], \quad (19)$$

где $K_{рч} = K_{ЭУС}K_{ЭУП}$ – результирующий коэффициент передачи напряжения электронных устройств передающей аппаратуры.

На выходе оптического сумматора (ОС) с коэффициентом передачи мощности $K_{P,ОС}$ действует оптическое поле с напряжённостью

$$E_{ОС}(t) = \sqrt{K_{P,ОС}} \cdot [E_{ИМЦ1}(t) + E_{ОФВ}(t)].$$

Для распределения ключа используется оптический аттенюатор (ОА) с коэффициентом передачи мощности $K_{P,ОА}$, который обеспечивает на выходе оптической передающей станции оптическое поле с напряжённостью

$$E_{ОПс}(t) = \sqrt{K_{P,ОА}} \cdot E_{ОС}(t) = \sqrt{K_{P,ОА}K_{P,ОС}} \cdot [E_{ИМЦ1}(t) + E_{ОФВ}(t)].$$

С учётом (17) и (19) находим

$$E_{\text{опс}}(t) = -j \cdot \sqrt{K_{P,OA} K_{P,OC} K_{P,ИМЦ}} \cdot \frac{E_{\text{ОКГ}}(t)}{2} \cdot \exp \left[j\pi K_{PЧ} \frac{U_{RF}}{U_{\pi}} [I_k \cos(\theta_{RF}) - Q_k \sin(\theta_{RF})] \right] + \rightarrow \\ \leftarrow + \sqrt{K_{P,OA} K_{P,OC} K_{P,ИМЦ} K_{P,ОФВ}} \cdot \frac{E_{\text{ОКГ}}(t)}{2} \cdot \exp \left[j\pi K_{PЧ} \frac{U_{PЧ}}{U_{\pi}} [I_k \cos(\theta_{PЧ}) + Q_k \sin(\theta_{PЧ})] \right].$$

Для упрощения записи формул введём обозначения для коэффициента фазовой модуляции радиосигнала

$$\alpha_{PЧ} = \pi K_{PЧ} U_{PЧ} / U_{\pi}. \quad (20)$$

и для параметров модуляции

$$\begin{aligned} \alpha_1 &= \alpha_{PЧ} [I_k \cos(\theta_{PЧ}) - Q_k \sin(\theta_{PЧ})], \\ \alpha_2 &= \alpha_{PЧ} [I_k \cos(\theta_{PЧ}) + Q_k \sin(\theta_{PЧ})]. \end{aligned} \quad (21)$$

На практике коэффициент передачи оптических фазовращателей близок к единице. Так, например, оптоволоконные вращатели фазы ОМ-FPS вносят потери мощности до 0,1 дБ на рабочих длинах волн 780, 1060, 1310 и 1550 нм. Это позволяет в дальнейшем принять $K_{P,ОФВ} = 1$. Тогда

$$E_{\text{опс}}(t) = \frac{\sqrt{K_{P,OA} K_{P,OC} K_{P,ИМЦ}}}{2} \cdot E_{\text{ОКГ}}(t) \cdot \{-j \cdot \exp(j\alpha_1) + \exp(j\alpha_2)\}. \quad (22)$$

Выводы. Представлена структура оптического квадратурного модулятора на двух параллельно включённых интерферометрах Маха-Цендера с постоянными напряжениями смещения на всех плечах управляющих электродов для формирования оптического излучения с модуляцией фазоманипулированным радиосигналом на поднесущей частоте. Применение технологии когерентной оптической связи позволит распространять ключи с высокой скоростью на большие расстояния. Применение квадратурной фазовой манипуляции (QPSK) обеспечивает помехоустойчивость к внешним воздействиям при высокой скорости передачи данных. Преимущества каждого из этих подходов реализуется в предлагаемой когерентной оптической системе распределения ключа. Дальнейшие исследования будут направлены на разработку структуры приёмной станции и организационных мер по обеспечению информационной безопасности формируемого ключа для симметричного шифрования потока данных.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Былина М.С., Глаголев С.Ф., Дюбов А.С. Сравнительный анализ методов энергетического и когерентного приёма цифровых информационных оптических сигналов. Ч. 1. Энергетический приём // Тр. учебных заведений связи. – 2017. – Т. 3, № 3. – С. 12-20.
2. Былина М.С., Глаголев С.Ф., Дюбов А.С. Сравнительный анализ методов энергетического и когерентного приёма цифровых информационных оптических сигналов. Ч. 2. Когерентный приём // Тр. учебных заведений связи. – 2017. – Т. 3, № 4. – С. 21-28.
3. Махмуд Х.А.М., Румянцев К.Е. Формирование однополосной квадратурной фазовой манипуляции радиосигналов на поднесущих частотах в когерентной оптической системе коммуникации // Известия ЮФУ. Технические науки. – 2022. – № 5. – С. 209-220. – DOI: 10.18522/2311-3103-2022-5-209-220.
4. Махмуд Х.А.М., Румянцев К.Е., Аль-Бегат А.Х.Ш. Анализ восходящего лазерного канала спутниковой коммуникации в условиях атмосферной турбулентности // Известия ЮФУ. Технические науки. – 2023. – № 4. – С. 174-191. – DOI: 10.18522/2311-3103-2023-4-174-191.
5. Скляр Бернард. Цифровая связь. Теоретические основы и практическое применение. Digital Communications: Fundamentals and Applications. – 2-е изд. – М.: Вильямс, 2007. – 1104 с. – ISBN 0-13-084788-7.
6. Величко М.А., Наний О.Е., Сусьян А.А. Новые форматы модуляции в оптических системах связи // Lightwave Russian Edition. – 2005. – № 4. – С. 21-30.
7. Топчий П.Н., Елишев В.В., Румянцев Д.В. Применение квадратурной амплитудной манипуляции для повышения скорости передачи цифровой информации // 79-я научно-техническая конференция Санкт-Петербургского НТО РЭС им. А.С. Попова. – 2024. – С. 89-91.

8. *Феер К.* Беспроводная цифровая связь. Методы модуляции и расширения спектра. *Wireless Digital Communications: Modulation and Spread Spectrum Applications*. – М.: Радио и связь, 2000. – 552 с. – ISBN 5-256-01444-7.
9. *Hui R., Zhu B., Huang R., Allen C., Demarest K. and Richards D.* Subcarrier multiplexing for high-speed optical transmission // *IEEE Journal of Lightwave Technology*. – 2002. – Vol. 20, No. 3. – P. 417-427.
10. *Махмуд Х.А.М., Румянцев К.Е.* Спектральный анализ системы передачи с мультиплексированием оптических поднесущих, использующей формат модуляции QPSK // *Фундаментальные и прикладные аспекты компьютерных технологий и информационной безопасности: Матер. VII Всероссийской научно-технической конференции*. – 2021. – С. 77-81.
11. *Махмуд Х.А.М., Румянцев К.Е.* Моделирование мультиплексирования поднесущих Найквиста для оптической системы передачи // *II Всероссийская научно-практическая конференция «Digital Era»*. – 2022. – С. 77-83.
12. *Hui R., Zhu B., Huang R., Allen C., Demarest K., Richards D.* 10-Gb/s SCM fiber system using optical SSB modulation // *IEEE Photonics Technology Letters*. – 2001. – Vol. 13, No. 8. – P. 896-898.
13. *Афанасьев В.М.* Электрооптический модулятор по схеме интерферометра Маха-Цендера // *Прикладная фотоника*. – 2016. – Т. 3, № 4. – С. 341-344.
14. *Древко Д.Р., Зюрюкин Ю.А., Ушаков Н.М.* Модификации электрооптического модулятора Маха-Цендера для управления лазерным излучением повышенной мощности // *Наноэлектроника, нанофотоника и нелинейная физика: Матер. V конференции молодых учёных*. – Саратов, 2010. – С. 18-19.
15. *Вобликов Е.Д., Волынцев А.Б., Журавлев А.А., Кичанов А.В., Пономарев Р.С., Шевцов Д.И.* Интегрально-оптический модулятор на основе интерферометра Маха-Цендера с асимметричной топологией волноводов // *Тр. МАИ*. – 2011. – № 46.
16. *Пономарев Р.С., Вобликов Е.Д.* Некоторые вопросы работы интегрально-оптических модуляторов интенсивности // *Вестник Пермского ун-та. Сер.: Физика*. – 2011. – Вып. 2 (17). – С. 65-68.
17. *Кузьминов Ю.С.* Электрооптический и нелинейно-оптический кристалл ниобата лития. – М.: Наука, 1987. – 264 с.
18. *Зюрюкин Ю.А., Павлова М.В., Древко Д.Р.* Волновые уравнения для описания эффекта Поккельса в кристаллах и их анализ на примере кристалла ниобата лития // *Известие вузов. Прикладная нелинейная динамика*. – 2010. – Т. 8, № 5. – С. 125-137.
19. *Величко М.А., Наний О.Е., Сусьян А.А.* Новые форматы модуляции в оптических системах связи // *Lightwave Russian Edition*. – 2005. – № 4. – С. 21-30.
20. *Афанасьев В.М.* Электрооптический модулятор по схеме интерферометра Маха-Цендера // *Прикладная фотоника*. – 2016. – Т. 3, № 4. – С. 341-369.
21. *Mahmood H.A., Rumyantsev K.Y.* Effect of FBG Compensated Dispersion on SCM/ASK Radio over Fiber System // *12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. – IEEE, 2019. – P. 1-5.
22. *Махмуд Х.А.М., Румянцев К.Е.* Формирование однополосной квадратурной фазовой манипуляции радиосигналов на поднесущих частотах в когерентной оптической системе коммуникации // *Известия ЮФУ. Технические науки*. – 2022. – № 5. – С. 209-220. – DOI: 10.18522/2311-3103-2022-5-209-220.

REFERENCES

1. *Bylina M.S., Glagolev S.F., Dyubov A.S.* Sravnitel'nyy analiz metodov energeticheskogo i kogerentnogo priema tsifrovyykh informatsionnykh opticheskikh signalov. Ch. 1. Energeticheskiy priem [Comparative analysis of methods of energy and coherent reception of digital information optical signals. Part 1. Energy reception], *Tr. uchebnykh zavedeniy svyazi* [Transactions of educational institutions of communication], 2017, Vol. 3, No. 3, pp. 12-20.
2. *Bylina M.S., Glagolev S.F., Dyubov A.S.* Sravnitel'nyy analiz metodov energeticheskogo i kogerentnogo priema tsifrovyykh informatsionnykh opticheskikh signalov. Ch. 2. Kogerentnyy priem [Comparative analysis of methods of energy and coherent reception of digital information optical signals. Part 2. Coherent reception], *Tr. uchebnykh zavedeniy svyazi* [Transactions of educational institutions of communication], 2017, Vol. 3, No. 4, pp. 21-28.
3. *Makhmud Kh.A.M., Rumyantsev K.E.* Formirovaniye odnopolosnoy kvadraturnoy fazovoy manipulyatsii radiosignalov na podnesushchikh chastotakh v kogerentnoy opticheskoy sisteme kommunikatsii [Formation of single-sideband quadrature phase-shift keying of radio signals at subcarrier frequencies in a coherent optical communication system], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2022, No. 5, pp. 209-220. DOI: 10.18522/2311-3103-2022-5-209-220.

4. *Makhmud Kh.A.M., Rumyantsev K.E., Al'-Begat A.Kh.Sh.* Analiz voskhodyashchego lazernogo kanala sputnikovoy kommunikatsii v usloviyakh atmosfery turbulentsii [Analysis of the upward laser channel of satellite communication under atmospheric turbulence], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2023, No. 4, pp. 174-191. DOI: 10.18522/2311-3103-2023-4-174-191.
5. *Sklyar Bernard.* Tsifrovaya svyaz'. Teoreticheskie osnovy i prakticheskoe primeneniye. Digital Communications: Fundamentals and Applications [Digital Communications: Fundamentals and Applications]. 2nd ed. Moscow: Vil'yams, 2007, 1104 p. ISBN 0-13-084788-7.
6. *Velichko M.A., Naniy O.E., Sus'yan A.A.* Novye formaty modulyatsii v opticheskikh sistemakh svyazi [New modulation formats in optical communication systems], *Lightwave Russian Edition*, 2005, No. 4, pp. 21-30.
7. *Topchiy P.N., Elishev V.V., Rumyantsev D.V.* Primeneniye kvadrurnoy amplitudnoy manipulyatsii dlya povysheniya skorosti peredachi tsifrovoy informatsii [Application of quadrature amplitude shift keying to increase the digital information transmission rate], *79-ya nauchno-tekhnicheskaya konferentsiya Sankt-Peterburgskogo NTO RES im. A.S. Popova* [79th scientific and technical conference of the St. Petersburg Scientific and Technical Association of Radio Electronic Systems named after A.S. Popov], 2024, pp. 89-91.
8. *Feer K.* Besprovodnaya tsifrovaya svyaz'. Metody modulyatsii i rasshireniya spektra. Wireless Digital Communications: Modulation and Spread Spectrum Applications [Wireless Digital Communications: Modulation and Spread Spectrum Applications]. Moscow: Radio i svyaz', 2000, 552 p. ISBN 5-256-01444-7.
9. *Hui R., Zhu B., Huang R., Allen C., Demarest K. and Richards D.* Subcarrier multiplexing for high-speed optical transmission, *IEEE Journal of Lightwave Technology*, 2002, Vol. 20, No. 3, pp. 417-427.
10. *Makhmud Kh.A.M., Rumyantsev K.E.* Spektral'nyy analiz sistemy peredachi s mul'tipleksirovaniem opticheskikh podnesushchikh, ispol'zuyushchey format modulyatsii QPSK [Spectral analysis of a transmission system with multiplexing of optical subcarriers using the QPSK modulation format], *Fundamental'nye i prikladnye aspekty komp'yuternykh tekhnologiy i informatsionnoy bezopasnosti: Mater. VII Vserossiyskoy nauchno-tekhnicheskoy konferentsii* [Fundamental and applied aspects of computer technology and information security: Proceedings of the VII All-Russian scientific and technical conference], 2021, pp. 77-81.
11. *Makhmud Kh.A.M., Rumyantsev K.E.* Modelirovaniye mul'tipleksirovaniya podnesushchikh Naykvista dlya opticheskoy sistemy peredachi [Modeling of Nyquist subcarrier multiplexing for an optical transmission system], *II Vserossiyskaya nauchno-prakticheskaya konferentsiya «Digital Era»* [II All-Russian scientific and practical conference "Digital Era"], 2022, pp. 77-83.
12. *Hui R., Zhu B., Huang R., Allen C., Demarest K., Richards D.* 10-Gb/s SCM fiber system using optical SSB modulation, *IEEE Photonics Technology Letters*, 2001, Vol. 13, No. 8, pp. 896-898.
13. *Afanas'ev V.M.* Elektroopticheskiy modulyator po skheme interferometra Makha-Tsendera [Electro-optical modulator based on the Mach-Zehnder interferometer scheme], *Prikladnaya fotonika* [Applied Photonics], 2016, Vol. 3, No. 4, pp. 341-344.
14. *Drevko D.R., Zyuryukin Yu.A., Ushakov N.M.* Modifikatsii elektroopticheskogo modulyatora Makha-Tsendera dlya upravleniya lazernym izlucheniem povyshennoy moshchnosti [Modifications of the Mach-Zehnder electro-optical modulator for controlling high-power laser radiation], *Nanoelektronika, nanofotonika i nelineynaya fizika: Mater. V konferentsii molodykh uchenykh* [Nanoelectronics, nanophotonics and nonlinear physics: Proceedings of the V conference of young scientists]. Saratov, 2010, pp. 18-19.
15. *Voblikov E.D., Volyntsev A.B., Zhuravlev A.A., Kichanov A.V., Ponomarev R.S., Shevtsov D.I.* Integral'no-opticheskiy modulyator na osnove interferometra Makha-Tsendera s asimmetrichnoy topologiyey volnovodov [Integrated optical modulator based on a Mach-Zehnder interferometer with an asymmetric waveguide topology], *Tr. MAI* [Trudy MAI], 2011, No. 46.
16. *Ponomarev R.S., Voblikov E.D.* Nekotorye voprosy raboty integral'no-opticheskikh modulyatorov intensivnosti [Some aspects of operation of integrated-optical intensity modulators], *Vestnik Permskogo un-ta. Ser.: Fizika* [Bulletin of Perm University. Series: Physics], 2011, Issue 2 (17), pp. 65-68.
17. *Kuz'minov Yu.S.* Elektroopticheskiy i nelineynno-opticheskiy kristall niobata litiya [Electro-optical and nonlinear-optical crystal of lithium niobate]. Moscow: Nauka, 1987, 264 p.
18. *Zyuryukin Yu.A., Pavlova M.V., Drevko D.R.* Volnovye uravneniya dlya opisaniya effekta Pokkel'sa v kristallakh i ikh analiz na primere kristalla niobata litiya [Wave equations for describing the Pockels effect in crystals and their analysis using the example of a lithium niobate crystal], *Izvestie vuzov. Prikladnaya nelineynaya dinamika* [News of universities. Applied nonlinear dynamics], 2010, Vol. 8, No. 5, pp. 125-137.
19. *Velichko M.A., Naniy O.E., Sus'yan A.A.* Novye formaty modulyatsii v opticheskikh sistemakh svyazi [New modulation formats in optical communication systems], *Lightwave Russian Edition*, 2005, No. 4, pp. 21-30.

20. Afanas'ev V.M. Elektroopticheskiy modulyator po skheme interferometra Makha-Tsendera [Electro-optical modulator based on the Mach-Zehnder interferometer scheme], *Prikladnaya fotonika* [Applied Photonics], 2016, Vol. 3, No. 4, pp. 341-369.
21. Mahmood H.A., Rumyantsev K.Y. Effect of FBG Compensated Dispersion on SCM/ASK Radio over Fiber System, *12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2019, pp. 1-5.
22. Makhmud Kh.A.M., Rumyantsev K.E. Formirovanie odnopolosnoy kvadraturnoy fazovoy manipulyatsii radiosignalov na podnesushchikh chastotakh v kogerentnoy opticheskoy sisteme kommunikatsii [Formation of single-sideband quadrature phase shift keying of radio signals at subcarrier frequencies in a coherent optical communication system], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2022, No. 5, pp. 209-220. DOI: 10.18522/2311-3103-2022-5-209-220.

Румянцев Константин Евгеньевич – Южный федеральный университет; e-mail: rke2004@mail.ru; г. Таганрог, Россия; тел.: +79281827209. кафедра информационной безопасности телекоммуникационных систем; зав. кафедрой; д.т.н.; профессор.

Цыцорин Дмитрий Александрович – Южный федеральный университет; e-mail: cicorinda@mail.ru; г. Таганрог, Россия; тел.: +79381543393; военный учебный центр; начальник цикла – старший преподаватель.

Rumyantsev Konstantin Yevgenievich – Southern Federal University; e-mail: rke2004@mail.ru; Taganrog, Russia; phone: +79281827209; the Department of Information Security of Telecommunication Systems; head of department; dr. of eng. sc.; professor.

Tsytsorin Dmitry Alexandrovich – Southern Federal University; e-mail: cicorinda@mail.ru; Taganrog, Russia; phone: +79381543393; Military Training Center; head of the cycle – senior teacher.

УДК 681.5

DOI 10.18522/2311-3103-2026-1-77-86

А.Ю. Будко, Т.А. Гайда, З.А. Понимаш

ОЦЕНКА ПРОСТРАНСТВЕННОГО ПОЛОЖЕНИЯ БОРТОВОЙ КАМЕРЫ ПУТЕМ СОПОСТАВЛЕНИЯ АЭРОФОТОСНИМКОВ И ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ

Описан метод оценки пространственного положения бортовой камеры летательного аппарата, заключающийся в сопоставлении аэрофотоснимков и геопривязанных данных дистанционного зондирования Земли (ДДЗ) путем обнаружения нейросетевым детектором устойчивых в пространстве и времени реперных точек в обоих наборах данных и дальнейшим решением известной задачи *Perspective-n-Point (PnP)* для оценки матриц поворота и перемещения, минимизирующих ошибку перепроецирования, исходя из соответствий точек 3D мирового пространства и 2D точек их проекций на матрицу бортовой камеры, что может быть использовано для решения актуальной задачи локализации летательных аппаратов в условиях отсутствия сигналов глобальной навигационной спутниковой системы. В качестве устойчивых реперных точек, хорошо выделяемых на данных ДДЗ и аэрофотоснимках, в работе выбраны перекрестки дорог, альтернативным выбором могут являться другие локальные семантические паттерны изображений, характерные для конкретной местности. Поскольку прямое сопоставление снимков ДДЗ и бортовых снимков затруднено вследствие существенной разницы условий съемки, предлагается применение устойчивых детекторов реперных точек на основе алгоритмов искусственных нейронных сетей (ИНС). Для обучения устойчивого детектора создан смешанный датасет на основе спутниковых снимков и данных бортовой съемки. При разметке данных смешанного датасета использована функция нормализованного к единице 3D гауссиана с вершиной в центре перекрестка, график которой проецируется на 2D маску обучающего набора. Параметры функции гауссиана рассчитываются на основе радиуса окружности, описывающей перекресток. Использование функции нормализованного гауссиана с вершиной в геометрическом центре проекции перекрестка позволяет обучить сеть предсказывать вероятность принадлежности каждого пикселя изображения к перекрестку, с максимумом в центре перекрестка, что увеличивает точность позиционирования за счет более точной геопривязки реперной точки мирового 3D набора. В качестве детектора перекрестков в работе обучена ИНС архитектуры типа U-Net, в качестве метрики качества

обучения использован дифференцируемый аналог метрики *Dice*, в качестве оптимизатора использован *AdamW* в связке с косинусным планировщиком скорости обучения *CosineAnnealingLR*. В заключительной части работы приведены результаты сопоставления спутниковых данных и бортового снимка предложенным методом.

Локализация роботов; системы технического зрения; визуальная навигация; искусственные нейронные сети; ключевые точки; Perspective-n-Point.

A.Yu. Budko, T.A. Gayda, Z.A. Ponimash

ESTIMATION OF THE SPATIAL POSITION OF AN ON-BOARD CAMERA BY COMPARING AERIAL IMAGES AND SATELLITE IMAGE DATA

The article describes a method for estimating the spatial position of an onboard camera of an aircraft. This method involves comparing aerial photographs and georeferenced remote sensing (RSS) data by using a neural network detector to detect stable spatiotemporal reference points in both datasets. This method then solves the well-known Perspective-n-Point (PnP) problem for estimating rotation and translation matrices that minimize the reprojection error based on the correspondences between 3D world points and 2D points of their projections onto the onboard camera matrix. This approach can be used to solve the pressing problem of aircraft localization in the absence of global navigation satellite system signals. Road intersections are selected as stable spatiotemporal reference points that are clearly visible in RSS data and aerial photographs. Other local semantic image patterns characteristic of a particular area may serve as an alternative. Since direct comparison of remote sensing and airborne images is difficult due to significant differences in shooting conditions, the use of robust landmark detectors based on artificial neural network (ANN) algorithms is proposed. To train the robust detector, a mixed dataset was created using satellite and airborne imagery. The mixed dataset was labeled using a 3D Gaussian function normalized to unity with a apex at the intersection center, the graph of which is projected onto a 2D mask of the training set. The parameters of the Gaussian function are calculated based on the radius of the circle enclosing the intersection. Using a normalized 3D Gaussian function with a apex at the geometric center of the intersection projection allows the network to predict the probability of each image pixel belonging to the intersection, with a maximum at the intersection center, which increases positioning accuracy due to more precise georeferencing of the landmark point in the global 3D dataset. A U-Net-type artificial neural network was trained as an intersection detector. A differentiable analog of the Dice metric was used as a training quality metric. AdamW, coupled with a CosineAnnealingLR cosine learning rate planner, was used as an optimizer. The final section of the paper presents the results of comparing satellite data and airborne imagery using the proposed method.

Camera pose estimation; robot localization; machine vision systems; visual navigation; artificial neural networks; keypoints; Perspective-n-Point.

Введение. Задача оценки положения бортовой камеры (в зарубежной литературе, – «camera pose estimation», прим. авторов) тесно связана с задачами локализации и навигации роботов по визуальным данным. Наиболее актуальной областью применения данного класса задач является локализация и навигация беспилотных летательных аппаратов (БПЛА) в условиях отсутствия сигналов ГНСС, картографирование, разведка местности, охрана и мониторинг оперативной обстановки. Разновидность методов и подходов к визуальной навигации, основанных на сопоставлении информации с датчиков и электронной карты на борту робота получили название обзорно-сравнительных или корреляционно-экстремальных [1]. В данной работе рассматривается система локализации корреляционно-экстремального типа, основанная сопоставлении геопривязанных спутниковых снимков и данных бортовой съемки в видимом диапазоне.

Традиционные алгоритмы (корреляция, SIFT, SURF) хорошо изучены, но имеют фундаментальные ограничения по устойчивости к изменениям условий съемки и требуют больших вычислительных ресурсов для обработки детальных карт. Анализ работ показывает смещение фокуса современных исследований в сторону методов искусственного интеллекта, которые радикально улучшают точность и надежность традиционных методов [2–9]. Сверточные нейронные сети (CNN) [10–13] и автоэнкодеры используются для сжатия изображений до компактных «дескрипторов» местности, что резко снижает объем хранимых данных и ускоряет поиск соответствий. Архитектуры типа PoseNet [14] учатся напрямую определять координаты и ориентацию БАС по одному кадру, минуя традиционные этапы поиска ключевых точек. В целом, нейросети лучше справляются с изменениями освещения, сезона, ракурса и погодных условий по сравнению с классическими

алгоритмами (SIFT, SURF). Среди передовых решений можно выделить следующие несколько архитектур. Сиамские сети специализированы для сравнения двух изображений и определения степени их схожести, идеальны для задач поиска эталона. Трансформеры (ViT) и гибриды CNN-Transformer показывают превосходство в задачах анализа изображений, включая локализацию и детекцию объектов, благодаря механизму внимания, улавливающему глобальный контекст сцены [15, 16]. Сети для плотного сопоставления (Neighbourhood Consensus Networks) обеспечивают точное, пиксельное соответствие между изображениями даже в слаботекстурированных сценах.

Формальная постановка задачи. Допустим имеется два набора данных к сопоставлению: карта местности на основе данных дистанционного зондирования Земли (ДДЗ) и снимки с бортовых камер летательного аппарата. Прямое сопоставление затруднено вследствие различных условий съемки, оборудования, погодных условий и прочих факторов. Для решения задачи предлагается введение промежуточного экстрактора опорных точек, обеспечивающего надежное извлечение координат и параметров таких точек из обоих наборов данных. Таким образом, после обработки опорной карты получается облако точек 3D (карта) в мировых координатах и соответствующий некоторой области облака (карты), 2D набор точек (шаблон) проекций локальной области карты на матрицу камеры с некоторыми (аффинными, сдвиговыми, нелинейными) искажениями в локальных координатах, а решение задачи сопоставления двух наборов данных переводится в область корреляционно-экстремального поиска соответствия между двумя облаками точек с минимизацией ошибки трансформации.

Таким образом задачу можно разбить на подзадачи: обнаружение ключевых точек, сопоставление облаков ключевых точек, минимизация ошибки трансформации между наборами точек и оценка положения камеры наблюдателя.

Подзадача обнаружения ключевых точек. Вначале была исследована возможность применения классических методов извлечения и сопоставления по ключевым точкам на базе дескрипторов. В качестве тестового набора изображений взяты фрагмент подложки на базе спутниковых снимков и ортофото на основе снимков БпЛА одного и того-же участка местности. В ходе исследования апробированы следующие алгоритмы поиска ключевых точек и дескрипторного описания: Harris Corner Detector; FAST (Features from Accelerated Segment Test); SIFT (Scale-Invariant Feature Transform); SURF (Speeded-Up Robust Features); ORB (Oriented FAST and Rotated BRIEF). Для сопоставления дескрипторов исследованы методы: Brute-Force Matcher (BFMatcher); FLANN (Fast Library for Approximate Nearest Neighbors); KNN Matching (k-Nearest Neighbors); SNN (Shared Nearest Neighbors). Наилучших результатов в прямом сопоставлении спутниковых снимков и снимков с бортовых камер дескрипторными методами удалось добиться, используя связку SSN + SOSNet (Second Order Similarity Regularization for Local Descriptor Learning), рис. 1.

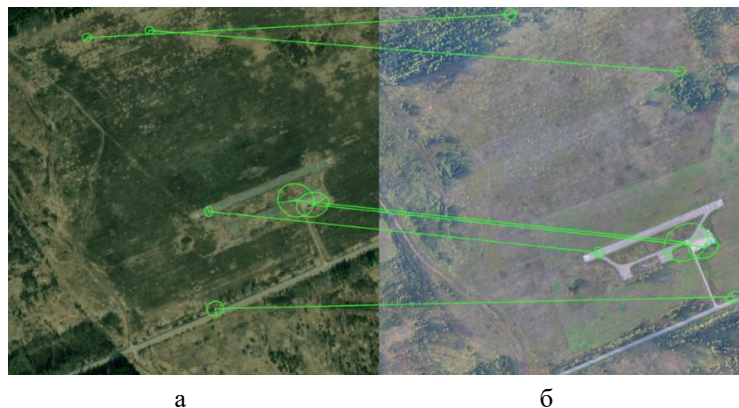


Рис. 1. Результаты сопоставления спутниковых снимков и снимков с бортовых камер на основе связки методов SSN + SOSNet: а – фрагмент подложки на базе спутниковых снимков, б – ортофото на основе снимков с борта БпЛА

Анализ рис. 1 показывает, что даже при исключении геометрических и масштабных различий путем составления из бортовых снимков ортофотоснимка, дескрипторные методы не могут обеспечить надежного извлечения пар точек ввиду, в частности на рис. 1 произведено правильное сопоставление 3 пары из 6 пар, т.е. 50%. Таким образом, результаты экспериментов по сопоставлению изображений и поиску фрагментов на подложке показали, что при существенной разнице в разрешении и уровне детализации изображений корректное сопоставление классическими методами машинного зрения без использования дополнительных априорных данных (например, приблизительных координат расположения фрагмента) или вспомогательных контекстных элементов вызывает существенные затруднения.

Для решения данной проблемы принято решение использовать в качестве опорных точек на спутниковых снимках и изображениях с бортовых камер перекрестки дорог, как достаточно устойчивые во времени и пространстве паттерны местности. В качестве детектора перекрестков был разработан алгоритм на базе ИНС архитектуры U-net [17–19]. Для обучения детектора перекрестков был создан смешанный датасет на основе спутниковых снимков и данных бортовой съемки БПЛА. Разметка произведена нормализованными гауссианами с вершиной в центре перекрестка, рис. 2.

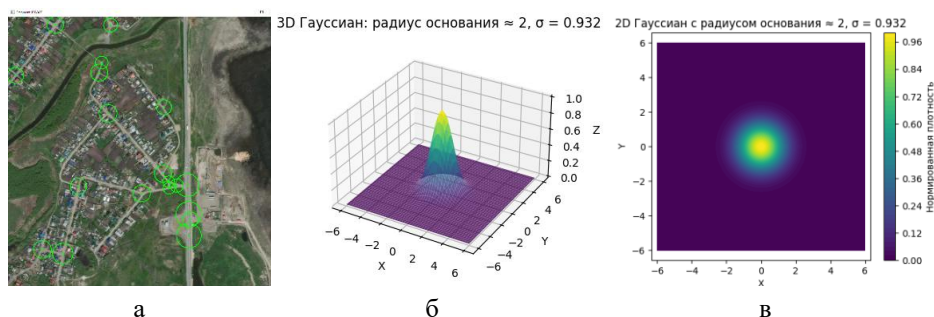
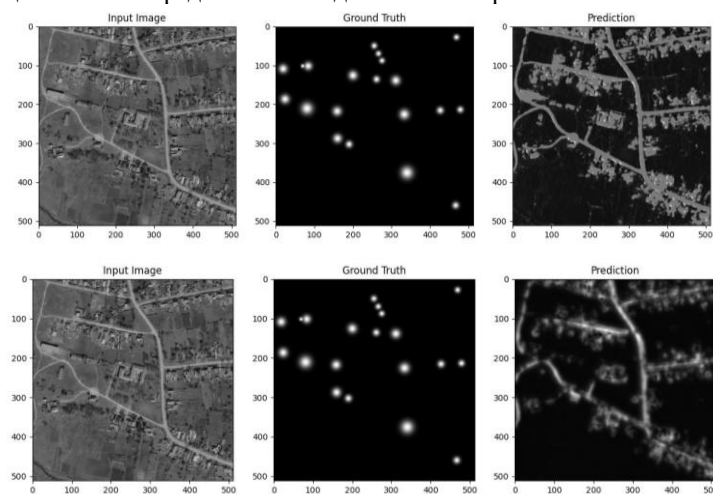


Рис. 2. Разметка датасета: а – фрагмент карты с размеченными перекрестками; б – 3D нормализованный гауссиан и его 2D проекция – в

Благодаря разметке с использованием функции нормализованных 3D гауссиан с вершиной в геометрическом центре проекции перекрестка сеть учится предсказывать вероятность принадлежности каждого пикселя к перекрестку, с максимумом в центре перекрестка, что увеличивает точность позиционирования. Далее была реализована архитектура U-Net на базе фреймворка torch, в качестве метрики качества разработан и реализован дифференцируемый аналог метрики Dice, в качестве оптимизатора использован AdamW в связке с косинусным планировщиком скорости обучения CosineAnnealingLR. На рис. 3 представлена наглядная эволюция качества предсказаний модели U-Net на различных эпохах.



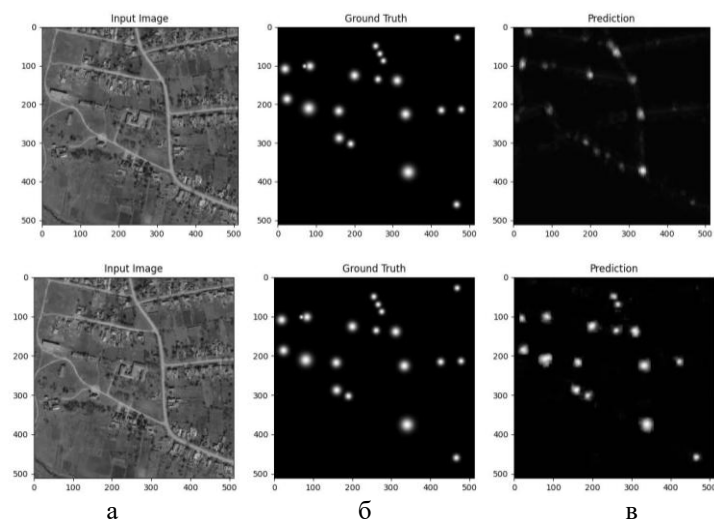


Рис. 3. Эволюция качества предсказаний модели UNet при обучении:
а – входное изображение, б – маска (учитель), в – результат работы ИНС

Примеры работы обученного детектора перекрестков на базе архитектуры UNet на данных из различных источников показаны на рис. 4.

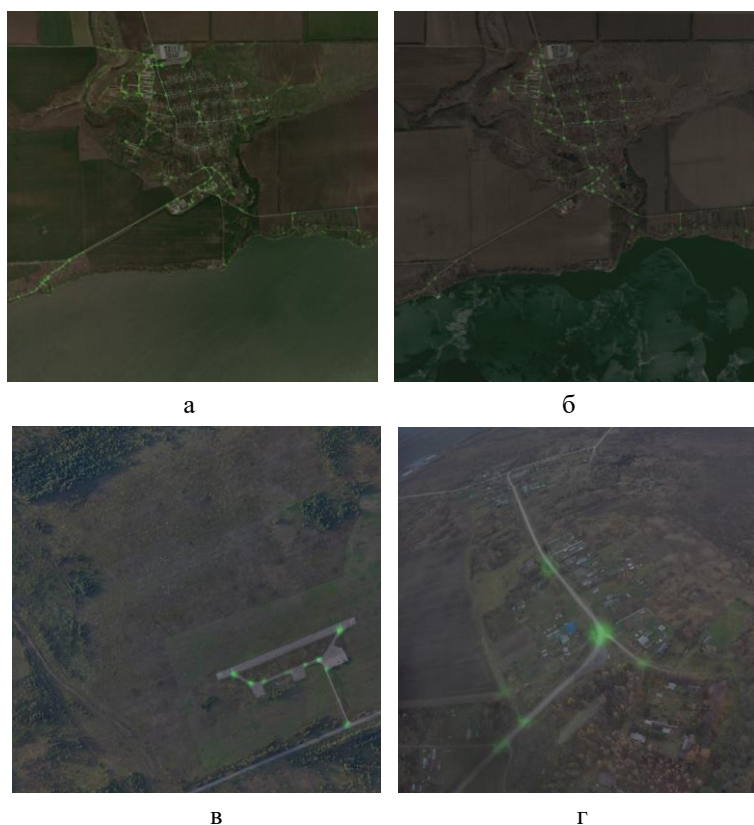


Рис. 4. Примеры работы обученного детектора перекрестков на базе архитектуры U-Net на данных из различных источников: (а) спутниковые снимки ESRI и (б) BING область 4x4 км, ортофото собранное из бортовых снимков (в) на области 2x2 км и сырой бортовой снимок со значительными перспективными искажениями (г)

Анализ рис. 4 показывает, что детектор обрабатывает на снимках из различных источников данных с различными масштабами изображения.

Подзадачи сопоставления облаков ключевых точек, минимизации ошибки трансформации между наборами точек и оценки положения камеры наблюдателя. Чтобы определить местоположение камеры в мировых координатах, нужно выполнить следующие шаги:

1. Выполнить предварительные процедуры:
 - ◆ произвести калибровку камеру (получить внутренние параметры – матрицы камеры и коэффициентов искажения);
 - ◆ получить набор известных точек в мировых координатах и соответствующих им точек на изображении (2D-3D пары).
2. Произвести решение задачи solvePnP, – solvePnP вычисляет вектор поворота ($rvec$) и вектор перемещения ($tvec$), которые показывают положение и ориентацию объекта в системе координат камеры на основе следующих аргументов:
 - ◆ набор 3D точек в мировых координатах ($objectPoints$);
 - ◆ набор 2D точек в изображении ($imagePoints$);
 - ◆ проекционная матрица камеры ($cameraMatrix$);
 - ◆ коэффициенты искажения ($distCoeffs$).
3. Произвести расчет позиции камеры в мировых координатах: $rvec$ и $tvec$ описывают координаты системы мировых точек в системе координат камеры, чтобы получить позицию камеры в мировых координатах, нужно инвертировать это преобразование:
 - ◆ получить из вектора поворота $rvec$ матрицу поворота R с помощью преобразования Родригеса;
 - ◆ рассчитать позицию камеры $X, Y, Z = -R^T * tvec$.

Задача вычисления позы Perspective-n-Point (PnP) pose computation [20] заключается в решении задачи поворота и перемещения, минимизирующей ошибку перепроецирования, исходя из соответствий точек 3D и 2D. Наборы точек для сопоставления и связь координатных систем при решении задачи PnP показаны на рис. 5.

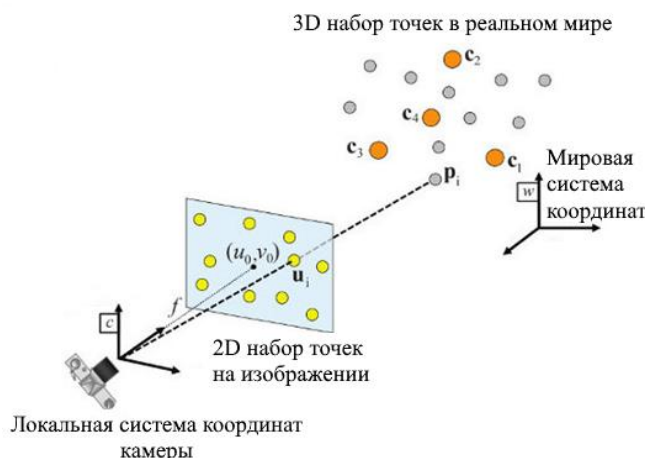


Рис. 5. Наборы точек для сопоставления и связь координатных систем при решении задачи PnP

Например, функция solvePnP библиотеки Open CV и связанные с ней функции оценивают позу объекта по набору точек объекта, соответствующим им проекциям изображения, а также собственной матрице камеры и коэффициентам искажения, рис. 6.

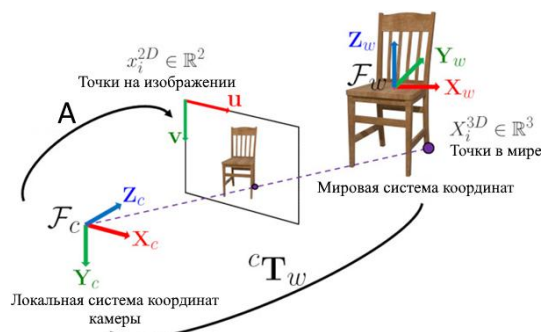


Рис. 6. Связь координатных систем при решении задачи PnP: ось X кадра камеры направлена вправо, ось Y – вниз, а ось Z – вперед

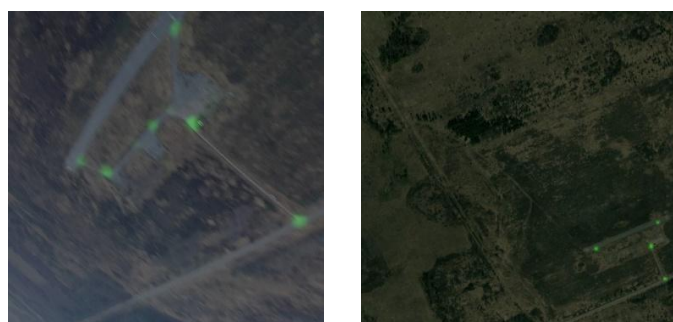
Точки, выраженные в мировой системе координат F_w , проецируются на плоскость изображения $[u, v]$ с использованием модели перспективной проекции и матрицы внутренних параметров камеры (1):

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \quad (1)$$

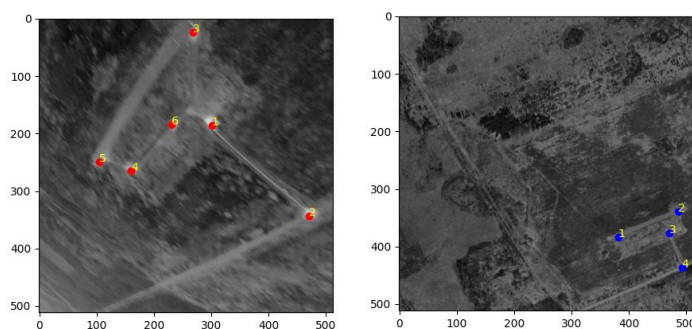
где f_x и f_y – эквивалентные фокусные расстояния по осям; c_x и c_y – пиксельные координаты центра кадра (точка на которую проецируется оптическая ось системы), r_{vec} и t_{vec} – векторы вращения и смещения.

Таким образом, расчетное положение описывается векторами вращения (r_{vec}) и смещения (t_{vec}), которые позволяют преобразовать трехмерную точку, выраженную в мировой системе координат, в кадр камеры.

Пример сопоставления данных бортовой камеры и данных ДДЗ предложенным методом показан на рис. 7.



а



б

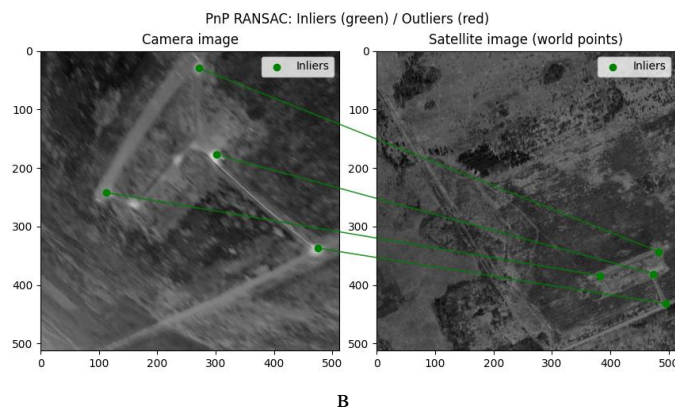


Рис. 7. Пример сопоставления данных бортовой камеры и данных ДДЗ:
 а – результаты работы детектора перекрестков, б – обнаруженные ключевые точки,
 в – результат работы алгоритма сопоставления

Анализ рис. 7 показывает, что предложенный подход позволяет уверенно сопоставить пару тестовых изображений с бортовой камеры даже с сильными проекционными и масштабными искажениями (левый ряд изображений) и фрагмента спутниковой карты (правый ряд изображений) при решении задачи PnP. Сопоставление и фильтрация ключевых точек произведены методом решения задачи PnP с оптимизацией RANSAC [21].

Заключение. В статье показан метод к оценки пространственного положения бортовой камеры летательного аппарата, основанный на выделении ключевых точек на спутниковой подложке и изображениях с бортовой камеры с помощью нейросетевого детектора и дальнейшим сопоставлением облаков точек с решением задачи PnP. В качестве детектора ключевых точек предложена архитектура архитектуры U-Net, обучаемая на смешанном датасете данных спутниковой и бортовой съемки с вероятностным пространственным распределением значений пикселей масок на основе нормализованных гауссиан, параметризуемых размерами окружности, описывающей область ключевой точки на изображении. Рассмотрены вопросы подготовки датасета и показаны примеры результатов работы нейросетевого детектора перекрестков и алгоритма сопоставления ключевых точек.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Степанов О.А., Носов А.С. Алгоритм коррекции навигационной системы по данным карты и измерителя, не требующий предварительного оценивания значений поля вдоль пройденной траектории // Гирокоспия и навигация. – 2020. – Т. 28, № 2 (109). – С. 70-90.
2. Scaramuzza D., Fraundorfer F. Visual Odometry [A Survey] // IEEE Robotics and Automation Magazine. – 2011.
3. Reichenbach M., Damker H., Federrath H., and Rannenber K. Individual Management of Personal Reachability in Mobile Communication // Proc. of the 13th International Information Security Conference, Copenhagen, Denmark, May 1997. – P. 164-174.
4. Lowe D.G. Distinctive image features from scale invariant keypoints // International journal of computer vision. – 2004. – 60 (2). – P. 91-110.
5. Bay H., Ess A., Tuytelaars T., and Van Gool L. Speeded-up robust features (SURF) // Computer vision and image understanding. – 2008. – 110 (3). – P. 346-359.
6. Cadena C., Carlone L., Carrillo H., Latif Y., Scaramuzza D., Neira J., Reid I.D., Leonard J.J. Simultaneous Localization and Mapping: Present, Future, and the Robust-Perception Age // IEEE Transactions on Robotics (cond. Accepted). – 2016.
7. Ghouaiel N. and Lefevre S. Coupling ground-level panoramas and aerial imagery for change detection // Geo-spatial Information Science. – 2016. – 19 (3). – P. 222-232.
8. Shukla P., Goel S., Singh P., and Lohani B. Automatic geolocation of targets tracked by aerial imaging platforms using satellite imagery // The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. – 2014. – 40 (1):381.

9. *Dunegan E.D., Luecke T.E.* Strapdown Inertial Navigation System Developments // *Proceedings of the AIAA Guidance and Control Conference.* – 1965.
10. *LeCun Y., Bottou L., Bengio Y., u Haffner P.* Gradient-based learning applied to document recognition // *Proceedings of the IEEE.* – 1998. – Vol. 86, No. 11. – P. 2278-2324.
11. *Krizhevsky A., Sutskever I., and Hinton G.E.* ImageNet Classification with Deep Convolutional Neural Networks // *Advances in Neural Information Processing Systems.* – 2012. – Vol. 25.
12. *Simonyan K. u Zisserman A.* Very Deep Convolutional Networks for Large-Scale Image Recognition // *arXiv preprint arXiv:1409.1556.* – 2015.
13. *He K., Zhang X., Ren S., u Sun J.* Deep Residual Learning for Image Recognition // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* – 2016. – P. 770-778.
14. *Kendall A., Grimes M., Cipolla R.* PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization // *Proceedings of the IEEE International Conference on Computer Vision (ICCV).* – 2015.
15. *Dosovitskiy A., Beyer L., Kolesnikov A. et al.* An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale // *arXiv preprint.* – 2020.
16. *Arandjelović R., Gronat P., Torii A., Pajdla T., & Sivic J.* NetVLAD: CNN architecture for weakly supervised place recognition // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* – 2016. – P. 5297-5307.
17. *Ronneberger O., Fischer P., Brox T.* U-Net: Convolutional Networks for Biomedical Image Segmentation / In: *Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) // Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science.* – Vol. 9351. – Springer, Cham, 2015. – https://doi.org/10.1007/978-3-319-24574-4_28.
18. *Ulmas Priit & Liiv Innar.* (2020). Segmentation of Satellite Imagery using U-Net Models for Land Cover Classification // *10.48550/arXiv.2003.02899.*
19. *Yadavendra S., Chand S.* Semantic segmentation and detection of satellite objects using U-Net model of deep learning // *Multimed Tools Appl.* – 2022. – 81. – P. 44291-44310. – <https://doi.org/10.1007/s11042-022-12892-2>.
20. *Eric Marchand, Hideaki Uchiyama, and Fabien Spindler.* Pose Estimation for Augmented Reality: A Hands-On Survey // *IEEE Transactions on Visualization and Computer Graphics.* – December 2016. – 22 (12). – P. 2633-2651.
21. *Martin A. Fischler and Robert C. Bolles.* Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography (англ.) // *Comm. of the ACM: journal.* – 1981. – June (Vol. 24). – P. 381-395. – doi: 10.1145/358669.358692.

REFERENCES

1. *Stepanov O.A., Nosov A.S.* Algoritm korrektsii navigatsionnoy sistemy po dannym karty i izmeritelya, ne trebuyushchiy predvaritel'nogo otsenivaniya znacheniy polya vdol' proydennoy traektorii [Algorithm for correcting a navigation system based on map and meter data that does not require preliminary estimation of field values along the trajectory], *Giroskopiya i navigatsiya* [Gyroscopy and Navigation], 2020, Vol. 28, No. 2 (109), pp. 70-90.
2. *Scaramuzza D., Fraundorfer F.* Visual Odometry [A Survey], *IEEE Robotics and Automation Magazine*, 2011.
3. *Reichenbach M., Damker H., Federrath H., and Rannenberg K.* Individual management of personal reachability in mobile communication, *Proc. of the 13th International Information Security Conference, Copenhagen, Denmark, May 1997*, pp. 164-174.
4. *Lowe D.G.* Distinctive image features from scale invariant keypoints, *International journal of computer vision*, 2004, 60 (2), pp. 91-110.
5. *Bay H., Ess A., Tuytelaars T., and Van Gool L.* Speeded-up robust features (SURF), *Computer vision and image under-standing*, 2008, 110 (3), pp. 346-359.
6. *Cadena C., Carlone L., Carrillo H., Latif Y., Scaramuzza D., Neira J., Reid I.D., Leonard J.J.* Simultaneous Localization and Mapping: Present, Future, and the Robust-Perception Age, *IEEE Transactions on Robotics (cond. Accepted)*, 2016.
7. *Ghouaiel N. and Lefevre S.* Coupling ground-level panoramas and aerial imagery for change detection, *Geo-spatial Information Science*, 2016, 19 (3), pp. 222-232.
8. *Shukla P., Goel S., Singh P., and Lohani B.* Automatic geolocation of targets tracked by aerial imaging platforms using satellite imagery, *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2014, 40 (1):381.
9. *Dunegan E.D., Luecke T.E.* Strapdown Inertial Navigation System Developments, *Proceedings of the AIAA Guidance and Control Conference*, 1965.
10. *LeCun Y., Bottou L., Bengio Y., u Haffner P.* Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 1998, Vol. 86, No. 11, pp. 2278-2324.

11. Krizhevsky A., Sutskever I., and Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems*, 2012, Vol. 25.
12. Simonyan K. u Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition, *arXiv preprint arXiv:1409.1556*, 2015.
13. He K., Zhang X., Ren S., u Sun J. Deep Residual Learning for Image Recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778.
14. Kendall A., Grimes M., Cipolla R. PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.
15. Dosovitskiy A., Beyer L., Kolesnikov A. et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, *arXiv preprint*, 2020.
16. Arandjelović R., Gronat P., Torii A., Pajdla T., & Sivic J. NetVLAD: CNN architecture for weakly supervised place recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5297-5307.
17. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation / In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science*, Vol. 9351. Springer, Cham, 2015. Available at: https://doi.org/10.1007/978-3-319-24574-4_28.
18. Ulmas Priit & Liiv Innar. (2020). Segmentation of Satellite Imagery using U-Net Models for Land Cover Classification, *10.48550/arXiv.2003.02899*.
19. Yadavendra S., Chand S. Semantic segmentation and detection of satellite objects using U-Net model of deep learning, *Multimed Tools Appl.*, 2022, 81, pp. 44291-44310. Available at: <https://doi.org/10.1007/s11042-022-12892-2>.
20. Eric Marchand, Hideaki Uchiyama, and Fabien Spindler. Pose Estimation for Augmented Reality: A Hands-On Survey, *IEEE Transactions on Visualization and Computer Graphics*, December 2016, 22 (12), pp. 2633-2651.
21. Martin A. Fischler and Robert C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Comm. of the ACM: journal*, 1981, June (Vol. 24), pp. 381-395. doi: 10.1145/358669.358692.

Будко Артем Юрьевич – Южный федеральный университет; e-mail: abudko@sfedu.ru; г. Таганрог, Россия; тел.: +78634371694; с.н.с. НИИ РИПУ; доцент.

Гайда Татьяна Александровна – Южный федеральный университет; e-mail: gaidatat@mail.ru; г. Таганрог, Россия; инженер.

Понимаш Захар Алексеевич – Южный федеральный университет; e-mail: ponimashz@mail.ru; г. Таганрог, Россия; тел.: +79885464739; аспирант; генеральный директор ООО «ФракталТех».

Budko Artem Yuryevich – Southern Federal University; e-mail: abudko@sfedu.ru; Taganrog, Russia; senior researcher.

Gaida Tatyana Aleksandrovna – Southern Federal University; e-mail: gaidatat@mail.ru; Taganrog, Russia; engineer.

Ponimash Zahar Alekseevich – Southern Federal University; e-mail: ponimashz@mail.ru; Taganrog, Russia; phone: +79885464739; graduate student; general director of FractalTech LLC.

УДК 621.396

DOI 10.18522/2311-3103-2026-1-86-96

М.В. Грачев, Ю.Н. Паршин

ПРОПУСКНАЯ СПОСОБНОСТЬ МОБИЛЬНОЙ МИМО СИСТЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ В НЕОДНОРОДНОЙ СРЕДЕ

Рассматривается излучение и прием электромагнитной волны в сплошной неоднородной среде подвижным источником и приемником. На основе лучевой теории проведен расчет коэффициента передачи сигнала от источника к приемнику в зависимости от их положения в пространстве и параметров неоднородности среды. Совокупность коэффициентов передачи для нескольких передающих и приемных пунктов составляет матрицу канальных коэффициентов МИМО системы передачи информации. Проведен спектральный анализ канальных коэффициентов в зависимости от параметров движения передающих и приемных пунктов. Установлено отличие спектра

в однородной и неоднородной средах. Полученные спектры существенно отличаются от классического спектра Джейкса, что подчеркивает необходимость учета неоднородной структуры среды при моделировании современных ММО систем. Исследуются флуктуации канальных коэффициентов, обусловленные пространственной неоднородностью среды, а также влияние скорости и направления движения передатчика и приемника на спектральные характеристики канала. Проведен расчет эргодической пропускной способности ММО системы передачи информации при движении в неоднородной сплошной среде. Показано, что сложная структура среды формирует непостоянное распределение амплитуд и фаз многолучевых сигналов, что приводит к немономонным изменениям пропускной способности. Установлено, что максимальная пропускная способность достигается при совпадении направления движения с направлением центра пучка лучей. Показано, что при ортогональной ориентации вектора скорости к пучку лучей модуль скорости движения практически не влияет на характеристики канала. Результаты исследования позволяют более точно учитывать физический характер среды и динамику движения источников, что важно для проектирования адаптивных методов обработки сигналов и оптимизации ММО систем нового поколения. Представленные зависимости могут быть использованы для повышения надежности передачи данных и максимизации пропускной способности в условиях мобильности передатчика сигнала в неоднородной сплошной среде.

Сплошная неоднородная среда; движущийся источник излучения; ММО система передачи информации; пропускная способность.

M.V. Grachev, Yu.N. Parshin

A CAPACITY OF MOBILE MIMO COMMUNICATION SYSTEM IN INHOMOGENEOUS MEDIUM

The propagation and reception of electromagnetic waves in a continuous, heterogeneous medium by mobile sources and receivers is considered. Based on ray theory, the calculation of signal transmission coefficients from the source to the receiver, depending on their positions in space and parameters of medium heterogeneity, is carried out. A set of transmission coefficients for multiple transmitting and receiving points constitute a matrix of MIMO channel coefficients for an information transmission system. Spectral analysis of channel coefficients, depending on motion parameters of transmitting and receiving points, has been performed. The difference in spectrum in homogeneous and heterogeneous media has been determined. The obtained spectra significantly differ from the classical Jakes spectrum, emphasizing the need to account for heterogeneous structure of medium in modeling of modern MIMO systems. Fluctuations in channel coefficients due to spatial heterogeneity of the medium have been investigated, as has the influence of the speed and direction of movement of the transmitter and receiver on the spectral properties of the channel. When moving in a heterogeneous continuous medium, the ergodic capacity of a MIMO information transmission system has been calculated. It has been demonstrated that the complex nature of the medium leads to an unstable distribution of amplitudes and phases for multipath signals, resulting in non-monotonic variations in bandwidth. Maximum capacity has been found to occur when the direction of movement coincides with that of the beam center. It has also been shown that when the velocity vector is orthogonal to the beam direction, the velocity magnitude has a negligible effect on channel properties. The findings of the study allow for a more accurate consideration of the physical characteristics of the environment and the dynamic behavior of sources, which is crucial for developing adaptive signal processing techniques and optimizing MIMO systems of the next generation. The provided dependencies can be utilized to enhance the reliability of data transmission and maximize the throughput in situations where the signal transmitter is mobile within an inhomogeneous and continuous environment.

Continuous inhomogeneous medium; moving radiation source; MIMO information transmission system; capacity.

Введение. Повышение требований к эффективности систем передачи информации стимулировало исследования в области более полного учета характеристик среды распространения радиоволн. Результатом проведенных исследований является создание технологии ММО [1], учитывающей преобразование сигнала при его распространении от каждой передающей антенны к каждой приемной антенне. При этом для описания свойств среды распространения используются различные модели, основанные на предположении об однородности среды. В работах [2, 3] исследуются характеристики канальных матриц с использованием геометрических и статистических моделей. Как пра-

вило, геометрическое моделирование предполагает, что среда распространения однородная, а многолучевость является следствием локальных неоднородностей. Различные модели среды распространения приведены также в стандарте COST 259.

Вместе с тем воздействие температуры, примесей различной концентрации и других факторов изменяют характер распространения сигнала, что влияет на характеристики ММО системы передачи информации. В работах [4, 5] исследуется пропускная способность неподвижной ММО системы в неоднородной сплошной среде, характеризуемой градиентом диэлектрической проницаемости. Установлено, что неоднородность среды вызывает многолучевой характер распространения сигнала, что необходимо учитывать при создании алгоритмов пространственного кодирования и декодирования. Неоднородность среды может быть существенным фактором работы в системах передачи информации IoT [6, 7].

Известны работы, в которых рассматривается влияние неоднородностей среды на характеристики ММО-канала при статических условиях. В работе [8] проведено моделирование ММО системы передачи информации в двухмерных неоднородных средах с помощью численных методов. Исследование эффекта Доплера в неоднородной среде показало, что неоднородности приводят к возникновению множественных путей распространения, каждый из которых вносит собственное смещение частоты [8, 10]. В работах [11–14] отмечено, что в неоднородной среде эффект Доплера имеет особенности, а смещение частоты зависит не только от радиальной скорости взаимного движения источника и приемника. Для ММО систем это означает, что при движении источника или приёмника флуктуации канального коэффициента будут формировать широкий спектр доплеровских флуктуаций, отличающийся от классических моделей однородной среды. В известных информационных источниках рассматривается влияние неоднородности среды в контексте беспроводных сенсорных сетей, где условия передачи могут значительно варьироваться на малых масштабах. Тем не менее, систематического учёта таких неоднородностей при оценке пропускной способности мобильных ММО систем в большинстве исследований пока нет. Основные работы по предмету касаются либо однородных моделей канала, либо статических сценариев с заданной геометрией среды.

Проводимое в данной работе исследование направлено на анализ характера флуктуаций канальных коэффициентов в неоднородной сплошной среде, а также расчет пропускной способности ММО системы передачи информации при движении передающей части системы. В отличие от ранее существовавших подходов, учитывается неоднородность среды непосредственно при моделировании канала ММО с движущимся передатчиком и анализом влияния скорости и направления движения на спектральный состав канальных флуктуаций и величину пропускной способности.

Целью данной работы является повышение пропускной способности ММО системы путем учета влияния неоднородности среды и характера движения мобильной системы передачи информации.

Постановка задачи. Для описания спектра многолучевого сигнала при его распространении от движущегося источника, где сигнал претерпевает искажения из-за множества отражений и рассеяний, используют модель Джейкса [15]. Классическое распределение Джейкса описывается следующим выражением [1] для спектральной плотности мощности сигнала:

$$S(f_d) = \begin{cases} \frac{A}{\sqrt{1 - (f_d/f_{d\max})^2}}, & |f_d| \leq f_{d\max}, \\ 0, & |f_d| > f_{d\max}, \end{cases}$$

где $f_{d\max} = f_0 v/c$ – максимальное смещение частоты вследствие эффекта Доплера, f_0 – частота сигнала, v – скорость движения объекта, c – скорость распространения ЭМВ (электромагнитной волны) в среде.

При анализе спектра принимаемого сигнала от движущегося передатчика необходимо учитывать ограничения: модель предполагает, что все пути распространения сигнала независимы и статистически одинаковы, что может не всегда быть верным при наличии сложных взаимодействий между лучами, например, фокусировки или усиления некоторых путей. Спектр Джейкса получен для однородной среды при условии, когда количество лучей многократно превышает единицу, а их влияние на общий сигнал можно рассматривать как случайный процесс с гауссовским распределением [16]. Таким образом, спектр Джейкса является важным инструментом для анализа радиосигналов в условиях многолучевого распространения с учетом доплеровских сдвигов частоты

В работе [4] установлено, что при распространении ЭМВ в неоднородной сплошной среде также наблюдается многолучевое распространение сигнала, обусловленное изменением электрических параметров среды. Сравнивая условия распространения ЭМВ в неоднородной среде и при расчете спектра Джейкса, можно отметить, что обе модели учитывают влияние многочисленных путей распространения ЭМВ с различными задержками и частотными сдвигами.

В работах [5, 6] с использованием лучевой теории распространения ЭМВ предложена модель неоднородной сплошной среды в виде слоистой структуры, которая состоит из плотно прилегающих слоев, разделенных плоскими границами. В этой структуре свойства среды изменяются вдоль одной из осей прямоугольной системы координат, оставаясь постоянными в плоскостях, перпендикулярных этой оси. При увеличении числа слоев и одновременно уменьшении толщины слоев такую среду распространения ЭМВ можно считать квазинепрерывно-слоистой.

Задача анализа распространения ЭМВ в данной структуре сводится к решению задач многократной дифракции на границах раздела двух сред. Так как для малой толщины слоя изменения свойств среды в слое минимальны, при моделировании амплитуды сигнала учитывается только преломленная составляющая луча в каждом слое, а отраженная составляющая учитывается только при выполнении условий полного внутреннего отражения [17, 18]. Фаза сигнала при прохождении границы раздела двух слоев не изменяется. Однако, при ее расчете учитывается геометрия траектории каждого луча, а также доплеровская составляющая фазы $2\pi f_d t$.

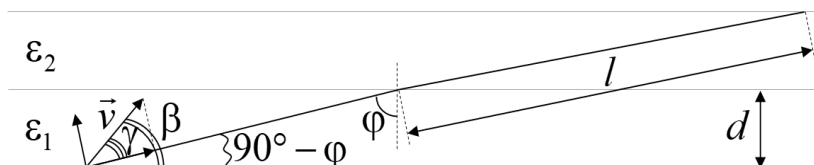


Рис. 1. Геометрия распространения сигнала от движущегося источника в неоднородной среде

На рис. 1 представлена геометрия траектории луча движущегося источника ЭМВ при прохождении границы раздела двух слоев. Угол γ между вектором скорости $\vec{v}$ и лучом можно определить через угол между вектором скорости и границей раздела сред β , а также угол падения на границу раздела двух сред φ :

$$\gamma = \beta + \varphi - \frac{\pi}{2}.$$

Эти углы влияют на характер распространения сигнала, так как при определенных условиях они могут изменять величину сдвига частоты вследствие эффекта Доплера. Радиальная составляющая вектора скорости определяется как проекция вектора скорости на линию распространения сигнала в рассматриваемом слое:

$$V_r = V \cos \gamma = V \cos \left(\beta + \varphi - \frac{\pi}{2} \right),$$

где V – модуль скорости движения источника ЭМВ.

Канальный коэффициент ММО системы, описывающий изменение сигнала движущегося источника при его распространении в среде, зависит от условий распространения сигнала, которые могут быть изменены различными внешними факторами, такими как плотность среды или наличие препятствий, а также сдвигом частоты, который зависит от скорости движения источника излучения. С учетом этих факторов выражение для одного луча, регистрируемого неподвижным приемником, имеет вид:

$$s_{RX}(t) = s_{TX}(t) H e^{-j2\pi f_d t},$$

где $f_d = \frac{f_0 V_r}{c_1}$ – смещение частоты сигнала вследствие эффекта Доплера, $s_{TX}(t)$ комплексная амплитуда излучаемого сигнала, H – канальный коэффициент, который зависит от свойств среды распространения, c_1 – скорость распространения ЭМВ в 1-м слое.

Если приемная антенна имеет апертуру конечных размеров, то выходной сигнал получается суммированием K лучей и определяется соотношением:

$$s_{RX}(t) = s_{TX}(t) H(t),$$

где $H(t) = \sum_{k=1}^K H_k e^{-j2\pi f_{dk} t}$ – коэффициент передачи сигнала, изменяющийся во времени,

$f_{dk} = \frac{f_0 V}{c_1} \cos \gamma_k$ – доплеровский сдвиг частоты сигнала k -го луча, γ_k – угловое направление k -го луча относительно вектора скорости V излучателя.

На рис. 2 показаны траектории распространения ЭМВ при движении источника излучения со скоростью 40 м/с при различных угловых направлениях в среде с градиентом диэлектрической проницаемости $6 \times 10^{-6} \text{ м}^{-1}$; частота сигнала равна 3 кГц. Стрелки на рисунке указывают направление движения передатчика для угловых направлений $\gamma = 3^\circ, 45^\circ, 83^\circ$. При этом угловая координата $\gamma = 3^\circ$ соответствует направлению центра прямой видимости приемной антенны из точки расположения передающей антенны. Направление $\gamma = 83^\circ$ соответствует направлению центра пучка лучей, излучаемых передающей антенной и попадающих на апертуру приемной антенны.

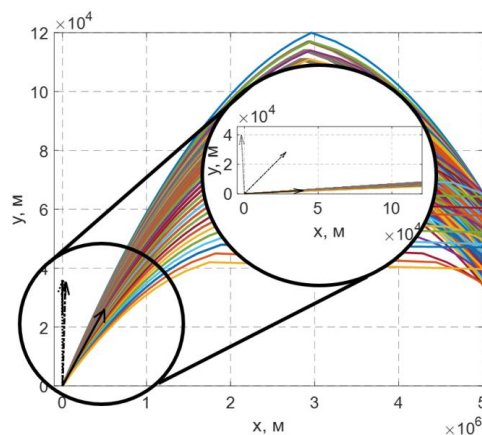


Рис. 2. Траектории ЭМВ и направление движения передатчика

Зависимость канального коэффициента ММО системы от скорости движения источника ЭМВ можно выразить через модуль скорости V и параметры среды, такие как диэлектрическая проницаемость, с учетом скорости доплеровского изменения фазы сигнала. Для модели с многослойным распространением сигнала фазовый сдвиг, возникающий вследствие движения источника, определяется как сумма сдвигов фаз, которые возникают в каждом из слоев среды [19, 20]. Каждая из этих фаз зависит от пути, который сигнал проходит через соответствующий слой среды. Этот сдвиг фазы зависит от параметров среды и скорости движения источника, что позволяет моделировать распространение ЭМВ в динамических условиях, таких как мобильные коммуникации или радиолокация.

Анализ пропускной способности ММО системы в неоднородной среде. В условиях неоднородной сплошной среды, где характеристики преломления и поглощения изменяются в пространстве, формируется сложная структура канала. Это оказывает прямое влияние на матрицу канальных коэффициентов ММО системы передачи информации и, следовательно, на максимально достижимую пропускную способность. Поэтому расчет пропускной способности в таких условиях требует учета суммарной амплитуды и фазы всех лучей, падающих на раскрыт приемных антенн, их волнового взаимодействия.

Необходимо отметить, что эргодическая пропускная способность рассчитывалась в результате усреднения во времени по реализациям канальных коэффициентов. Число значимых значений канальных коэффициентов, участвующих в усреднении пропускной способности, выбирается из условий некоррелированности этих значений на заданном интервале времени. Для заданной модели неоднородной сплошной среды число независимых значений канальных коэффициентов зависит от числа лучей на раскрыт приемной антенны. При проектировании систем беспроводной связи в неоднородных средах необходимо учитывать эти факторы для оптимизации характеристик передачи. Таким образом, пропускная способность ММО системы определяется путем усреднения пропускной способности по ансамблю реализаций канальных матриц

$$\mathbf{H}_m = \begin{bmatrix} H_{11}(m\Delta t) & H_{12}(m\Delta t) \\ H_{21}(m\Delta t) & H_{22}(m\Delta t) \end{bmatrix} \text{ для 2-х передающих антенн и 2-х приемных антенн. Реали-}$$

зации канальных матриц получаются путем взятия отсчетов во времени через интервал, равный интервалу корреляции канальной матрицы $\Delta t = \tau_c$. Пропускная способность определяется для гауссовского непрерывного канала в соответствии с выражением [1]

$$C = \frac{1}{M} \sum_{m=1}^M \log_2 \det \left(\mathbf{I} + \frac{q}{N_{TX}} \mathbf{H}_m^H \mathbf{H}_m \right),$$

где q – энергетический потенциал системы, N_{TX} – число передающих антенн, M – число реализаций канальной матрицы, используемых для усреднения, $\mathbf{H}_m$ – матрица коэффициентов передачи H между парами N_{TX} передающих антенн и N_{RX} приемных антенн. Реализации канальной матрицы формируются из зависимости $\mathbf{H}(t)$ в моменты времени, отстоящими на интервал времени больше, чем интервал корреляции.

Результаты моделирования. Моделирование канальных коэффициентов проводилось для ММО с числом антенн $N_{TX} = N_{RX} = 2$; частота сигнала $f_0 = 3$ кГц. Передающие антенны располагаются в точках с координатами $x_{TX1} = 0$, $y_{TX1} = 0$, $x_{TX2} = 0$, $y_{TX2} = 0,5\lambda$. Координаты приемной антенны $x_{RX1} = 50\lambda$, $y_{RX1} = 0$, $x_{RX2} = 50\lambda$, $y_{RX2} = 0,5\lambda$. Эффективная длина антенн равна $D = \frac{\lambda}{2\pi}$. Моделирование проводилось при следующих параметрах приемной системы и среды распространения: область моделирования по вертикальной оси ограничена интервалом $[0; 15\lambda]$ и разбита на 500 слоёв, диэлектрическая проницаемость среды изменяется в пределах $\varepsilon = 5, 1 \dots 5, 4$.

На рис. 3 приведены зависимости эргодической пропускной способности от скорости движения источника излучения для различных направлений движения передатчика γ ; энергетический потенциал системы равен $q=70$ дБ.

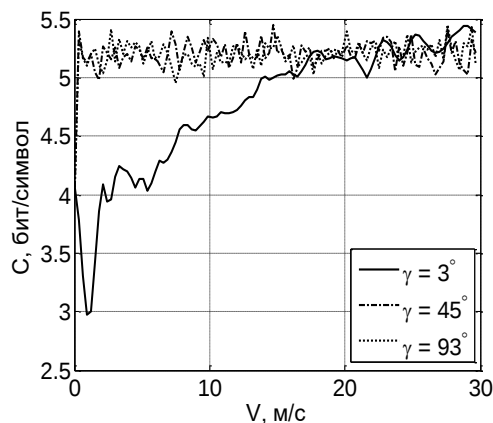


Рис. 3. Зависимость пропускной способности от скорости движения передатчика

Графики рис. 3 показывают, что влияние скорости движения на пропускную способность наиболее заметно при направлении движения ортогональном направлению середины углового сектора пучка лучей $\gamma = 3^\circ$. При увеличении скорости величина пропускной способности приближается к величине для направления вектора скорости, близкой к направлению середине углового сектора пучка лучей. В последнем случае величина скорости практически не влияет на пропускную способность ММО системы.

На рис. 4 показано, как изменяется пропускная способность канала связи в зависимости от направления движения передающей антенны при фиксированном значении энергетического потенциала системы и скорости движения равной $V=20$ м/с. Эффективность передачи информации неравномерна при различном направлении движения передатчика.

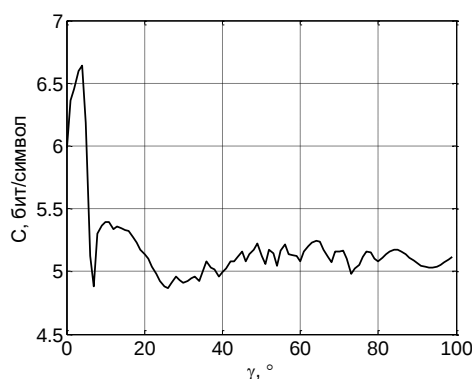


Рис. 4. Зависимость пропускной способности от направления движения передатчика

При изменении направления движения передатчика пропускная способность изменяется немонотонно в пределах 10...12 %. Наибольшее значение пропускной способности наблюдается при совпадении направления движения с направлением максимальной концентрации лучей (рис. 2). Эти колебания обусловлены чередованием конструктивной и деструктивной интерференции приходящих сигналов. Доплеровский эффект изменяется в зависимости от угла движения относительно приёмника, а неоднородная структура среды усиливает данные

колебания. Осцилляционный характер кривой на рис. 4 отражает сложное взаимодействие динамики доплеровских сдвигов, многолучевого распространения и пространственной неоднородности среды. Доплеровский эффект вносит регулярные частотные изменения, связанные со скоростью движения передающей антенны, а сама среда, обладающая неоднородной структурой, усиливает эти колебания за счёт непостоянного распределения отражённых и рассеянных сигналов. В совокупности это приводит к тому, что пропускная способность канала не изменяется монотонно, а демонстрирует выраженные пульсации, что имеет важное значение при проектировании систем беспроводной связи в условиях подвижных источников.

На рис. 5 приведена реализация во времени действительной части одного из канальных коэффициентов, отражающая его непериодическое и нерегулярное поведение. Зависимость была получена при $V = 40$ м/с, частоте сигнала 3 кГц, координатах передающей антенны $x_{TX} = y_{TX} = 0$, координатах приемной антенны $x_{RX} = 10\lambda$, $y_{RX} = 0,5\lambda$ и различных угловых направлениях вектора скорости движения передатчика.

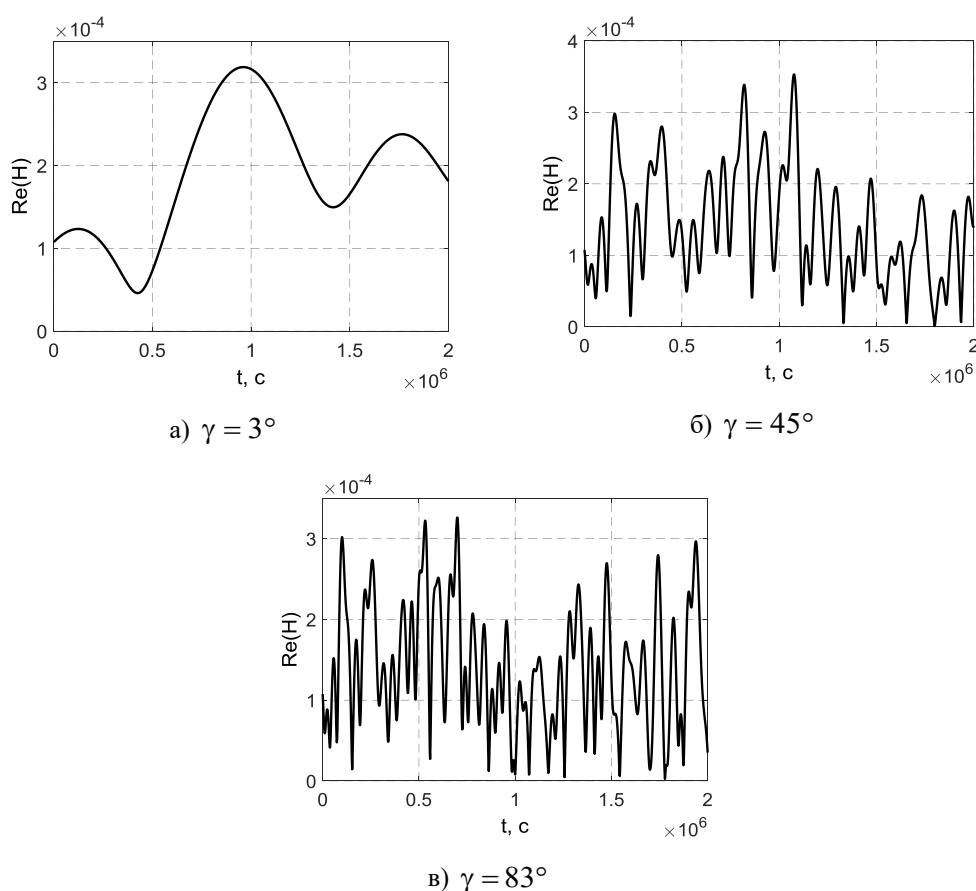


Рис. 5. Зависимость действительной части канального коэффициента от времени

Как следует из графиков рис. 5 частота флуктуаций канального коэффициента во многом определяется угловой координатой центра пучка лучей, попадающих на апертуру приемной антенны. При совпадении направления вектора скорости передатчика с направлением центра пучка лучей скорость флуктуаций наибольшая (рис. 5,в), а при ортогональной ориентации – наименьшая (рис. 5,а). Таким образом, установлено, что спектр зависит от параметров неоднородной среды, в чем принципиально отличается от спектра Джейкса [15].

Полученные зависимости канального коэффициента во времени подтверждают предположение, что канал является нестационарным и стохастическим, что важно учитывать при моделировании и разработке адаптивных методов обработки сигналов в системах беспроводной связи.

Заключение. Проведенный анализ канальных коэффициентов в мобильной ММО системе передачи информации показал влияние скорости движения источника и приемника электромагнитной волны на спектр флуктуаций канальных коэффициентов и пропускную способность ММО системы. Увеличение скорости движения приводит к расширению спектра флуктуаций канальных коэффициентов, что объясняется неоднородностью среды распространения. Соответственно пропускная способность также немонотонно зависит от скорости движения. Установлено, что влияние скорости в значительной мере зависит от взаимного расположения источника и приемника электромагнитной волны, а также от направления скоростей их взаимного движения. Новизна результатов работы заключается в том, что установлено критическое влияние взаимной ориентации и направления движения приемной и передающей антенн на эффективность передачи данных. Помимо обоснованного возрастания роли среды распространения, разработанная модель неоднородной сплошной среды позволяет провести более детальный учёт физического характера среды и механизма передачи, что важно при проектировании ММО систем передачи информации нового поколения.

Работа выполнена при поддержке гранта Российского научного фонда 24-29-00850, <https://rscf.ru/en/project/24-29-00850/> в Рязанском государственном радио-техническом университете им. В.Ф. Уткина.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Kuhn V. Wireless Communications over MIMO Channels. Applications to CDMA and Multiple Antenna Systems. – Chichester: John Wiley & Sons, Ltd., 2006. – 402 p.
2. Alcocer Ochoa A., Parra Michel R., Kontorovitch V.Ya. Geometrical Modeling of Wideband MIMO Channels // Computación y Sistemas. – 2006. – Vol. 9, No. 4. – P. 332-344.
3. Almers P., Bonek E., Burr A., et al. Survey of Channel and Radio Propagation Models for Wireless MIMO Systems // EURASIP Journal on Wireless Communications and Networking. – 2007. – Art. ID 19070. – 19 p. – DOI: 10.1155/2007/19070.
4. Parshin Yu., Grachev M. Modeling of the MIMO Communication System's Channel Matrix in Inhomogeneous Continuum // Proceedings of the 26th International Conference on Digital Signal Processing and its Applications (DSPA). – Moscow, Russian Federation, 2024. – P. 1-4.
5. Parshin Yu., Grachev M. Communication System's Channel Capacity in Inhomogeneous Continuum with a Given Spatial Structure // Proceedings of the 13th Mediterranean Conference on Embedded Computing (MECO). – Budva, Montenegro, 2024. – P. 1-5.
6. Parshin A., Parshin Yu. Investigation of Efficient Receiving of Ultra Low Power Signal for IoT Application // Proceedings of the 8th Mediterranean Conference on Embedded Computing (MECO). – Budva, Montenegro, 2019. – P. 32-35.
7. Parshin A., Parshin Yu. Adaptive Filtering of Non-Gaussian Flicker Noise // Proceedings of the 9th Mediterranean Conference on Embedded Computing (MECO). – Budva, Montenegro, 2020. – P. 529-533.
8. Yuan S.S.A., He Z., Sun S., Chen X., Huang C., Sha W.E.I. Electromagnetic Effective-Degree-of-Freedom Limit of a MIMO System in 2-D Inhomogeneous Environment // Electronics. – 2022. – Vol. 11, No. 19. – P. 3232. – DOI: 10.3390/electronics11193232.
9. Eden A. The Search for Christian Doppler. – Wien: Springer-Verlag, 1992. – 136 c. – DOI: 10.1007/978-3-7091-6677-2.
10. Weichman P.B. Doppler Effects in Heterogeneous Media with Applications to Ocean Acoustic Modeling // Physical Review E. – 2005. – Vol. 72, No. 6. – Art. 066602. – DOI: 10.1103/PhysRevE.72.066602.
11. Krylovich V.I. Nonstationary Doppler effect for waves propagating in nonhomogeneous media // Journal of Engineering Physics. – 1982. – Vol. 43, No. 4. – P. 1120-1125. – DOI: 10.1007/BF00827778.
12. Glushchenko A., Glushchenko E. The Doppler Effect in Non-Reciprocal Media // European Research. – 2017. – No. 33. – DOI: 10.20861/2410-2873-2017-33-003.
13. Gerlach C., Kromer A. Wave Propagation in Variable Media // Journal of Acoustic and Electromagnetic Studies. – 2018. – Vol. 41, No. 2. – P. 105-119.

14. Schaefer R., Thompson L., Berg J. Applications of Doppler Shift in Irregular Media // *Applied Physics Reports*. – 2021. – Vol. 54, No. 7. – P. 367-381.
15. Jakes W.C. (ed.). *Microwave Mobile Communications*. – New York: Wiley-Interscience, 1974. – 642 p.
16. Clarke R.H. A Statistical Theory of Mobile Radio Reception // *Bell System Technical Journal*. – 1968. – Vol. 47, No. 6. – P. 957-1000.
17. Семенов Д.И., Афанасьев С.А., Санников Д.Г. Основы теории распространения электромагнитных волн: учеб. пособие. – Ульяновск: УлГУ, 2012. – 112 с.
18. Гавриленко В.Г., Яинов В.А. Распространение электромагнитных волн в неоднородной плазме: учебное пособие. – Нижний Новгород: Нижегородский госуниверситет, 2015. – 101 с.
19. Никольский В.В., Никольская Т.И. Электродинамика и распространение радиоволн. – М.: Наука, 1989. – 560 с.
20. Гольдштейн Л.Д., Зернов Н.В. Электромагнитные поля и волны. – М.: Советское радио, 1971. – 448 с.

REFERENCES

1. Kuhn V. *Wireless Communications over MIMO Channels. Applications to CDMA and Multiple Antenna Systems*. Chichester: John Wiley & Sons, Ltd., 2006, 402 p.
2. Alcocer Ochoa A., Parra Michel R., Kontorovitch V.Ya. Geometrical Modeling of Wideband MIMO Channels, *Computación y Sistemas*, 2006, Vol. 9, No. 4, pp. 332-344.
3. Almers P., Bonek E., Burr A., et al. Survey of Channel and Radio Propagation Models for Wireless MIMO Systems, *EURASIP Journal on Wireless Communications and Networking*. 2007, Art. ID 19070, 19 p. DOI: 10.1155/2007/19070.
4. Parshin Yu., Grachev M. Modeling of the MIMO Communication System's Channel Matrix in Inhomogeneous Continuum, *Proceedings of the 26th International Conference on Digital Signal Processing and its Applications (DSPA)*. Moscow, Russian Federation, 2024, pp. 1-4.
5. Parshin Yu., Grachev M. Communication System's Channel Capacity in Inhomogeneous Continuum with a Given Spatial Structure, *Proceedings of the 13th Mediterranean Conference on Embedded Computing (MECO)*. Budva, Montenegro, 2024, pp. 1-5.
6. Parshin A., Parshin Yu. Investigation of Efficient Receiving of Ultra Low Power Signal for IoT Application, *Proceedings of the 8th Mediterranean Conference on Embedded Computing (MECO)*. Budva, Montenegro, 2019, pp. 32-35.
7. Parshin A., Parshin Yu. Adaptive Filtering of Non-Gaussian Flicker Noise, *Proceedings of the 9th Mediterranean Conference on Embedded Computing (MECO)*. Budva, Montenegro, 2020, pp. 529-533.
8. Yuan S.S.A., He Z., Sun S., Chen X., Huang C., Sha W.E.I. Electromagnetic Effective-Degree-of-Freedom Limit of a MIMO System in 2-D Inhomogeneous Environment, *Electronics*, 2022, Vol. 11, No. 19, pp. 3232. DOI: 10.3390/electronics11193232.
9. Eden A. *The Search for Christian Doppler*. Wien: Springer-Verlag, 1992, 136 c. DOI: 10.1007/978-3-7091-6677-2.
10. Weichman P.B. Doppler Effects in Heterogeneous Media with Applications to Ocean Acoustic Modeling, *Physical Review E*, 2005, Vol. 72, No. 6, Art. 066602. DOI: 10.1103/PhysRevE.72.066602.
11. Krylovich V.I. Nonstationary Doppler effect for waves propagating in nonhomogeneous media, *Journal of Engineering Physics*, 1982, Vol. 43, No. 4, pp. 1120-1125. DOI: 10.1007/BF00827778.
12. Glushchenko A., Glushchenko E. The Doppler Effect in Non-Reciprocal Media, *European Research*, 2017, No. 33. DOI: 10.20861/2410-2873-2017-33-003.
13. Gerlach C., Kromer A. Wave Propagation in Variable Media, *Journal of Acoustic and Electromagnetic Studies*, 2018, Vol. 41, No. 2, pp. 105-119.
14. Schaefer R., Thompson L., Berg J. Applications of Doppler Shift in Irregular Media, *Applied Physics Reports*, 2021, Vol. 54, No. 7, pp. 367-381.
15. Jakes W.C. (ed.). *Microwave Mobile Communications*. New York: Wiley-Interscience, 1974, 642 p.
16. Clarke R.H. A Statistical Theory of Mobile Radio Reception, *Bell System Technical Journal*, 1968, Vol. 47, No. 6, pp. 957-1000.
17. Sementsov D.I., Afanas'ev S.A., Sannikov D.G. *Osnovy teorii rasprostraneniya elektromagnitnykh voln: ucheb. posobie* [Fundamentals of electromagnetic wave propagation: textbook]. Ul'yanovsk: UlGU, 2012, 112 p.
18. Gavrilenko V.G., Yashnov V.A. *Rasprostraneniye elektromagnitnykh voln v neodnorodnoy plazme: ucheb. posobie* [Electromagnetic wave propagation in inhomogeneous plasma: textbook]. Nizhniy Novgorod: Nizhegorodskiy gosuniversitet, 2015, 101 p.
19. Nikol'skiy V.V., Nikol'skaya T.I. *Elektrodinamika i rasprostraneniye radiovoln* [Electrodynamics and radio wave propagation]. Moscow: Nauka, 1989, 560 p.
20. Gol'dshteyn L.D., Zernov N.V. *Elektromagnitnye polya i volny* [Electromagnetic fields and waves]. Moscow: Sovetskoe radio, 1971, 448 p.

Грачев Максим Викторович – Рязанский государственный радиотехнический университет им. В.Ф. Уткина; e-mail: grachev.m.v@rsreu.ru; г. Рязань, Россия; тел.: +79966162890; доцент; SPIN-код 2986-9109.

Паршин Юрий Николаевич – Рязанский государственный радиотехнический университет им. В.Ф. Уткина; e-mail: parshin.y.n@rsreu.ru; г. Рязань, Россия; тел.: +79105037576; зав. кафедрой; SPIN-код 9185-4280.

Grachev Maxim Victorovich – Ryazan State Radio Engineering University; e-mail: grachev.m.v@rsreu.ru; Ryazan, Russia; phone: +79966162890; associate professor; SPIN-code 2986-9109.

Parshin Yury Nikolaevich – Ryazan State Radio Engineering University; e-mail: parshin.y.n@rsreu.ru; Ryazan, Russia; phone: +79105037576; head of department; SPIN-code 9185-4280.

УДК 004.658

DOI 10.18522/2311-3103-2026-1-96-112

Д.Ю. Зорькин, А.А. Богнюков, Т.Е. Кожанова**РАЗРАБОТКА И ВНЕДРЕНИЕ КОРПОРАТИВНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ НА ПРОМЫШЛЕННОМ ПРЕДПРИЯТИИ ЗАВОДА «АвтоВАЗ»**

В условиях глобальной цифровизации промышленности разработка и внедрение корпоративных информационных систем (КИС) приобретает стратегическое значение для повышения операционной эффективности и конкурентоспособности предприятий. В представленном исследовании рассмотрен кейс интеграции ERP-системы «1С: Управление предприятием» на заводе «АвтоВАЗ» – ключевом игроке российской автомобильной промышленности. Цель работы заключается в разработке и внедрении корпоративной информационной системы (КИС) на базе платформы «1С: ERP», направленной на оптимизацию корпоративной информационной системы (КИС). Методологическая база исследования объединила аналитический, графический и сравнительный подходы, а также практическую апробацию решений в программной среде «1С». Основное внимание уделено проектированию алгоритмов управления производственными циклами, формированию ресурсных спецификаций и настройке сценариев планирования. В рамках исследования разработаны модели прогнозирования спроса, проведён анализ производственных мощностей и выполнена балансировка выпуска с учётом приоритетности моделей (Lada Granta, Vesta, Largus). Реализация системы позволила сократить время обработки заказов на 30%, минимизировать простои складов на 18–22% за счёт синхронизации логистических графиков, а также повысить точность контроля качества через интеграцию диагностических инструментов (CAN-bus, спектрофотометрия). Особый акцент сделан на преодолении институциональных и технологических барьеров, включая модернизацию устаревших методов планирования, обучение сотрудников работе с ERP-интерфейсами и внедрение гибридных облачных решений для обеспечения масштабируемости системы. Практическая значимость исследования подтверждена достижением прозрачности распределения ресурсов, снижением операционных издержек и формированием адаптивной производственной стратегии, ориентированной на динамику рынка. Результаты работы демонстрируют, что внедрение КИС не только оптимизирует текущие бизнес-процессы, но и создаёт основу для устойчивого развития предприятия в условиях цифровой трансформации. Полученный опыт может быть экстраполирован на другие машиностроительные и промышленные предприятия, сталкивающиеся с задачами автоматизации управления и интеграции данных в условиях высокой конкуренции. Перспективы дальнейших исследований связаны с анализом долгосрочных эффектов от внедрения ERP-систем, включая их влияние на инновационный потенциал и экосистему цепочки поставок.

Корпоративные информационные системы; ERP-системы; производственное планирование; цифровая трансформация; ресурсное управление; автомобильная промышленность; оптимизация бизнес-процессов; «1С:ERP»; «АвтоВАЗ».

D.Yu. Zorkin, A.A. Bognyukov, T.E. Kozhanova

DEVELOPMENT AND IMPLEMENTATION OF A CORPORATE INFORMATION SYSTEM AT THE AvtoVAZ INDUSTRIAL ENTERPRISE

In the context of global industrial digitalization, the development and implementation of corporate information systems (CIS) have become strategically critical for enhancing operational efficiency and competitiveness of enterprises. This study examines the integration case of the ERP system "1C: Enterprise Management" at the AvtoVAZ plant – a key player in the Russian automotive industry. The research aimed to optimize management and production processes through the automation of planning, resource accounting, and coordination of cross-functional interactions. The methodological framework combined analytical, graphical, and comparative approaches, as well as practical testing of solutions in the "1C" software environment. The focus was on designing algorithms for managing production cycles, forming resource specifications, and configuring planning scenarios. The study developed demand forecasting models, analyzed production capacities, and balanced output based on model prioritization (Lada Granta, Vesta, Largus). The system implementation reduced order processing time by 30%, minimized warehouse downtime by 18–22% through synchronized logistics schedules, and improved quality control accuracy via integrated diagnostic tools (CAN-bus, spectrophotometry). Special emphasis was placed on overcoming institutional and technological barriers, including modernizing outdated planning methods, training employees in ERP interfaces, and deploying hybrid cloud solutions to ensure system scalability. The practical significance of the research was confirmed by achieving resource allocation transparency, reducing operational costs, and forming an adaptive production strategy aligned with market dynamics. The results demonstrate that CIS implementation not only optimizes current business processes but also lays the foundation for sustainable enterprise development in the digital transformation era. The acquired experience can be extrapolated to other engineering and industrial enterprises facing challenges in management automation and data integration under competitive pressure. Future research prospects involve analyzing the long-term effects of ERP system adoption, including their impact on innovation potential and supply chain ecosystems.

Corporate information systems; ERP systems; production planning; digital transformation; resource management; automotive industry; business process optimization; "1C:ERP"; "AvtoVAZ".

Введение. Актуальность разработки и внедрения элементов корпоративной информационной системы (КИС) на предприятиях обусловлена стремительным развитием информационных технологий и их значением для повышения конкурентоспособности организаций [1]. Завод «АвтоВАЗ», являясь одним из крупнейших производителей автомобильной техники в России, остро нуждается в модернизации своих управленческих и производственных процессов для повышения эффективности работы и адаптации к условиям динамичного рынка [2].

Целью данной работы является разработка и внедрение элементов КИС, направленных на оптимизацию процессов управления, учета и планирования на предприятии [3]. Для достижения цели ставятся следующие задачи:

- ◆ исследование назначения и структуры КИС;
- ◆ анализ проблем, возникающих при внедрении информационных систем на предприятии [4];
- ◆ разработка элементов КИС для выбранных бизнес-процессов завода [5];
- ◆ обоснование и описание применяемых программных средств, включая систему «1С: ERP Управление предприятием».

Методы исследования включают аналитический, графический и сравнительный подходы, а также практическую реализацию предложений в программной среде «1С».

Структура работы состоит из теоретической части, где рассматриваются основы проектирования и внедрения КИС, и практической части, включающей разработку и реализацию элементов системы на примере завода «АвтоВАЗ». Теоретическая часть освещает назначение и ключевые характеристики ERP-систем, а также особенности современного рынка КИС. Практическая часть включает проектирование элементов корпоративной информационной системы, разработку алгоритмов производственных процессов, формирование ресурсных спецификаций, а также настройку сценариев планирования с

использованием инструментов ERP-системы. Кроме того, в рамках практической части выполняются прогнозирование спроса, анализ производственных мощностей и расчет экономических показателей.

Данное исследование способствует систематизации знаний о корпоративных информационных системах и их применении в условиях крупного промышленного предприятия.

Постановка задачи. В условиях цифровой трансформации промышленности предполагаемой проблемой завода «АвтоВАЗ» может являться неэффективность управления производственными и управленческими процессами, обусловленная:

- ◆ использованием устаревших методов планирования;
- ◆ несогласованностью данных между функциональными подразделениями;
- ◆ высокими операционными издержками из-за ручного учёта ресурсов;
- ◆ отсутствием прозрачности в распределении материалов и контроле качества.

Цель исследования – разработка и внедрение корпоративной информационной системы (КИС) на базе платформы «1C:ERP», направленной на оптимизацию бизнес-процессов, автоматизацию планирования и повышение конкурентоспособности предприятия.

Основные задачи:

1. Провести анализ текущих бизнес-процессов завода, включая снабжение, производство, сбыт и контроль качества, для выявления узких мест.
2. Спроектировать архитектуру КИС, интегрирующую модули ресурсного планирования, прогнозирования спроса и управления производственными циклами [3].
3. Реализовать алгоритмы балансировки выпуска продукции с учётом приоритетности моделей (Lada Granta, Vesta, Largus) и ограничений производственных мощностей.
4. Настроить сценарии автоматизированного документооборота, включая формирование спецификаций, заказов и логистических графиков.
5. Обеспечить интеграцию КИС с существующими ИТ-системами предприятия и роботизированными производственными линиями.
6. Разработать программу обучения сотрудников работе с ERP-интерфейсами и гибридными облачными решениями.
7. Оценить эффективность внедрения системы по метрикам:
 - ◆ сокращение времени обработки заказов;
 - ◆ снижение простоев складов;
 - ◆ повышение точности прогнозирования спроса;
 - ◆ улучшение контроля качества продукции.

Методы решения:

- ◆ Аналитический подход: сравнение данных до и после внедрения системы.
- ◆ Моделирование производственных процессов с использованием инструментов «1C:ERP».
- ◆ Статистический анализ данных (прогнозирование спроса, расчёт себестоимости).
- ◆ Практическая апробация решений на тестовых участках производства.

Ожидаемый результат: Создание адаптивной КИС, обеспечивающей сокращение времени обработки заказов на 30%, минимизацию логистических простоев на 18–22% и формирование единого цифрового контура для управления ресурсами в реальном времени.

Материалы и методы. В методологической базе были использованы аналитический, графический и сравнительный подходы [6]. В качестве примера можно рассмотреть крупное автомобильное предприятие «АвтоВАЗ» (АО «АвтоВАЗ»), специализирующееся на производстве легковых автомобилей и комплектующих к ним. Основная сфера деятельности – машиностроение, а именно разработка, сборка, реализация и сервисное обслуживание автомобильной техники на внутреннем и внешнем рынках.

Организация построена по дивизионно-функциональному принципу. Это означает, что предприятие разделено на несколько крупных блоков, которые могут иметь как функциональное (отдел снабжения, отдел сбыта, бухгалтерия, юридический отдел и т.д.), так и продуктовое деление (производственные цеха, подразделения по работе с дилер-

скими сетями, отделы испытаний и НИОКР). На верхнем уровне, как правило, формируются дирекции и департаменты, отвечающие за стратегическое планирование и принятие управленческих решений.

В рамках подобной структуры обычно выделяются следующие ключевые управленческие органы и подразделения:

1. Высшее руководство (правление, совет директоров). Определяет стратегические цели и задачи, контролирует финансовые результаты и инвестиционные проекты.
2. Функциональные департаменты (финансы, маркетинг, HR, ИТ, юридический и др.). Отвечают за методологию, регламенты, внутренний контроль, развитие своих направлений и поддержку основных производственных процессов.
3. Производственные подразделения (цеха, сборочные линии). Несут ответственность за непосредственный выпуск готовых изделий, соблюдение технологических нормативов, контроль качества и производительности.
4. Отделы продаж и снабжения. Фокусируются на внешних взаимодействиях: закупка ресурсов, логистика, реализация продукции, развитие дилерских и сервисных сетей.
5. Отдельные проектные группы или НИОКР-подразделения. Занимаются инновационными разработками, тестированием новых моделей и доводкой технических решений.

Для автомобильного завода наиболее характерны следующие отраслевые особенности:

1. Сбыт (реализация продукции).
 - ◆ Широкая дилерская сеть и развитая система автосалонов.
 - ◆ Акцент на маркетинговые кампании и продвижение новых моделей.
 - ◆ Необходимость постоянного анализа конкурентного рынка и учёта сезонных колебаний спроса.
2. Снабжение.
 - ◆ Закупка металла, пластика, комплектующих у различных отечественных и зарубежных поставщиков.
 - ◆ Жёсткие требования к качеству материалов, своевременности поставок и оптимизации складских запасов.
 - ◆ Многоуровневая логистика: от внешних поставщиков к собственным складским комплексам и далее к производственным конвейерам.
3. Учёт готовой продукции.
 - ◆ Организация складского учёта автомобилей и запасных частей на промежуточных складах и в дилерских центрах.
 - ◆ Отслеживание отгрузок и контроль остатков по складским площадям.
 - ◆ Взаимодействие с логистическими операторами и системой транспортировки готовой продукции.
4. Производство.
 - ◆ Многоступенчатый технологический процесс (штамповка кузовных деталей, сварка, окраска, сборка узлов, финальная сборка автомобилей).
 - ◆ Высокая автоматизация и роботизация производственных линий.
 - ◆ Постоянный мониторинг качества (тестовые стенды, испытательные полигоны, контроль соответствия стандартам).

Уровни планирования являются одним из важных факторов эффективной деятельности крупного промышленного предприятия, к которым относится и ПАО «АвтоВАЗ», является чёткая система планирования. На практике здесь выделяют три основных уровня планирования: стратегический, тактический и оперативный. Каждый из уровней выполняет собственные задачи и обеспечивает согласованную работу в рамках единой корпоративной структуры.

На стратегическом уровне решения принимают высшие органы управления, такие как совет директоров и топ-менеджмент компании. На данной ступени формируется общее видение развития «АвтоВАЗа» и определяются долгосрочные цели – например, выход на рынок принципиально новых моделей автомобилей, наращивание производственных мощностей или расширение ассортимента комплектующих. Стратегические решения

часто включают выбор приоритетных технологий, определение ключевых направлений инвестиций и постановку глобальных финансовых целей (например, рост рыночной доли или увеличение объёмов экспортных поставок).

К тактическому планированию относятся решения, принимаемые в профильных департаментах и службах (например, в департаменте производства, департаменте сбыта, отделе логистики). На этом уровне руководители определяют конкретные планы на год, квартал или месяц, ориентируясь на стратегические ориентиры. В частности, разрабатываются производственные программы, формируются бюджеты на закупку сырья и комплектующих, устанавливаются контрольные показатели (KPI) для основных подразделений, а также корректируются планы по маркетинговой активности или модернизации технологического оборудования. Тактический уровень обеспечивает согласованность между подразделениями и гибкое реагирование на текущую рыночную ситуацию, сохраняя при этом общий курс, заданный стратегией компании.

Оперативное планирование связано с повседневной деятельностью цехов, производственных линий и складских подразделений. Здесь ведётся точное расписание работ на неделю, день или смену, распределяются задания по конкретным рабочим местам и контролируется фактическое исполнение. Например, на этом уровне может идти речь о том, сколько автомобилей определённой модели должно быть выпущено в конкретную смену, как будет загружено оборудование и сколько материалов потребуется на ближайшие производственные циклы. Оперативные планы регулярно уточняются в режиме реального времени на основе данных, поступающих от автоматизированных систем, включая корпоративную информационную систему (КИС).

Результаты. В исследовании необходимо было разработать 3 вида номенклатуры разных типов. В системе 1C:ERP существует 5 типов номенклатуры [7]. Для проектирования ресурсных спецификаций применены методы статистического анализа данных [8], а также моделирование производственных циклов [9].

Товар – к этому типу относится вся «материальная» номенклатура, которая закупается, производится, реализуется предприятием и учитывается на складах. Возможны контроль остатков на складах, учет себестоимости, обеспечение потребностей и др.

Тара – номенклатура этого типа используется для учета многооборотной тары, которая может возвращаться поставщику товаров или передаваться на условиях возврата клиенту, покупающему товар в данной таре.

Работа – к этому типу относятся работы и услуги, которые могут закупаться у поставщиков и продаваться клиентам. Ключевым признаком номенклатуры данного типа является учет закупленных работ по количеству и себестоимости. При продаже клиентам работ, произведенных сторонними поставщиками, стоимость их закупки будет учитываться как себестоимость продажи. Работы учитываются аналогично товарам, но они не хранятся на складе, а числятся за подразделением, которое их произвело или закупило.

Услуга – номенклатура этого типа также может производиться, закупаться и продаваться. Ключевым отличием услуг от работ является отсутствие учета себестоимости. При закупках услуги сразу списываются на затраты. При продаже услуг себестоимость их оказания не рассчитывается.

Набор – комплекты, которые не хранятся на складе, а собираются в процессе отгрузки. Порядок работы с наборами описан в разделе «Динамические наборы».

В данном случае было выбрано 3 типа номенклатуры: товар, работа и услуга.

Если включен учет изделий, содержащих драгоценные металлы, алкогольной (спиртосодержащей) продукции, подконтрольной продукции животного происхождения, меховых изделий, табачной продукции или другой продукции, подлежащей обязательной маркировке, то тип номенклатуры дополнительно классифицируются по признаку принадлежности к этим товарным группам.

Схема последовательной передачи работ будет выглядеть следующим образом:

Закупка сырья и комплектующих -> Штамповка кузовных деталей -> Сварочные работы -> Покраска кузова -> Основная сборка автомобиля -> Установка двигателя и трансмиссии -> Установка салона и электрики -> Тестирование и контроль качества -> Отгрузка на склад готовой продукции -> Отгрузка дилерам.

В качестве примера следует рассмотреть производственное задание № PZ-AV-2024-01.

Текст задания:

«Дата формирования: 10 января 2024 года.

Срок исполнения: до 31 января 2024 года.

Ответственное подразделение: Цех сборки автомобилей (Основная линия сборки № 2).

Утверждено: Директором по производству

1. Наименование и объём производства

1. Продукция: Легковые автомобили LADA Vesta Sedan (код номенклатуры AVT-01-03).

2. Количество: 200 единиц (автомобилей).

2. Календарный план выполнения задания

1. Предварительная подготовка (закупка и комплектование узлов)

◆ Срок: 01.12.2024 – 04.12.2024.

◆ Ответственные: Отдел снабжения и склад материалов.

◆ Задачи:

■ Проверить наличие необходимых узлов (двигатели, трансмиссии, электрооборудование) и комплектующих (кузовные панели, колёса, стекло) на складе.

■ При нехватке – оперативно оформить заказы поставщикам/другим цехам.

Сборка кузова и сварочные операции

◆ Срок: 05.12.2024 – 17.12.2024.

◆ Ответственное подразделение: Сварочный цех.

◆ Задачи:

■ Подготовка кузовных панелей (штабель металла, штамповка деталей).

■ Сварка кузова в соответствии с технологической картой модели LADA Vesta Sedan.

Покраска кузова

◆ Срок: 18.12.2024 – 20.12.2024.

◆ Ответственное подразделение: Покрасочный цех.

◆ Задачи:

■ Грунтовка, окраска кузова, нанесение антикоррозионного покрытия.

■ Контроль качества и визуальный осмотр перед отправкой на сборочный конвейер.

Основная сборка (шасси, двигатель, салон и электрика)

◆ Срок: 21.12.2024 – 27.12.2024.

◆ Ответственное подразделение: Сборочный цех № 2 (Основная линия).

◆ Задачи:

■ Монтаж двигателя, трансмиссии, ходовой части.

■ Установка сидений, приборной панели, прокладка проводки, тестирование электроники.

■ Присоединение основных деталей кузова: капота, дверей, багажника.

Проверка и тестирование

◆ Срок: 28.12.2024 – 29.12.2024.

◆ Ответственные: Отдел технического контроля (ОТК).

◆ Задачи:

■ Пусконаладочные работы (проверка систем автомобиля на тестовом стенде).

■ Диагностика двигателя, тормозной и рулевой систем.

■ Выявление дефектов и устранение недочётов в случае их обнаружения.

Передача на склад готовой продукции

◆ Срок: 30.12.2024.

◆ Ответственные: Сборочный цех № 2 и склад готовой продукции.

- ♦ Задачи:
 - Оформление документов приёмки готовых автомобилей.
 - Регистрация в корпоративной информационной системе (КИС) с указанием фактического количества и серийных номеров.
 - Подготовка к отгрузке дилерским центрам.

Материалы и комплектующие (табл. 1):

Таблица 1

Материалы и комплектующие

Наименование / Код	Тип номенклатуры	Требуемое кол-во	Ед. изм.	Примечания
Кузовные детали Vesta / SP-05	Товар	200 комплектов	компл.	Поступают из штамповочного цеха
Двигатель 1,6 л / ENG-01-02	Товар	200 шт.	шт.	Собственные производственные мощности / склад
Комплект шин 16" / SP-04-12	Товар	200 комплектов	компл.	1 комплект = 4 шины
Покрасочные работы кузова / RB-01-07	Работа	200 кузовов	усл. ед.	В случае, если покраска частично выполняется сторонними подрядчиками
Предпродажная подготовка / SV-01-01	Услуга	200 авт.	усл. ед.	Обязательный этап перед складированием и отгрузкой

Прежде всего, следует отметить, что обеспечение качества продукции базируется на строгом соблюдении нормативов ГОСТ и внутренних регламентов завода «АвтоВАЗ». В частности, процедуры включают неразрушающий контроль сварных швов методом ультразвуковой дефектоскопии, оценку адгезии лакокрасочного покрытия по стандарту ISO 2409 и функциональное тестирование электрооборудования с использованием стендовых испытаний.

Кроме того, ключевым требованием является соблюдение правил техники безопасности. Так, обязательным условием выступает проведение целевого инструктажа для сотрудников, задействованных в сварочных, покрасочных и сборочных операциях, а также использование средств индивидуальной защиты (СИЗ) категории FFP3 и огнестойкой спецодежды.

По поводу логистики, сроки поставки комплектующих регламентируются договорённостями с отделом материально-технического обеспечения (МТО) и внешними контрагентами. В случае выявления дефицита ресурсов, ответственные лица обязаны немедленно уведомить руководителя проекта для активации резервных схем снабжения. При необходимости привлечения сторонних подрядчиков (например, для нанесения спецпокрытий), согласование осуществляется через МТО до начала работ.

Следует рассмотреть алгоритм производственного задания:

1. Инициализация проекта

На этапе получения задания ответственный руководитель анализирует техническое задание, включая сроки, объёмы выпуска и специфические требования (например, необходимость нестандартных материалов). Параллельно проводится верификация ресурсной базы с привлечением планово-диспетчерского отдела (ПДО) для минимизации рисков срыва графика.

2. Планирование ресурсов

Формирование спецификации материалов сопровождается аудитом складских остатков. При выявлении дефицита инициируются заявки поставщикам с резервированием производственных мощностей. Персонал инструктируется по технологическим картам, что обеспечивает синхронизацию этапов конвейера.

3. Технологические операции

◆ Подготовка сырья: Штамповка кузовных панелей выполняется на гидравлических прессах усилием 2500 тс с точностью позиционирования ± 0.1 мм.

◆ Сборка кузова: Роботизированная сварка по схеме jig-фреймов гарантирует геометрическую стабильность конструкции.

◆ Окраска: Нанесение трёхслойного покрытия в камерах с контролем влажности ($50 \pm 5\%$) и температуры ($23 \pm 2^\circ\text{C}$).

◆ Финальная сборка: Монтаж силового агрегата, ходовой части и интерьера с использованием torque-ключей, калиброванных под параметры крепежа.

4. Контроль качества

Промежуточные инспекции на каждом этапе включают:

◆ Визуальный осмотр сварных соединений (критерий: отсутствие пор и трещин).

◆ Измерение толщины ЛКП спектрофотометром X-Rite MA98.

◆ Тестирование электронных систем через диагностический интерфейс CAN-bus.

5. Документооборот и логистика

После успешной валидации оформляется пакет сопроводительной документации (паспорт изделия, сертификат соответствия ЕАЭС), а данные вносятся в корпоративную ERP-систему [10]. Отгрузка дилерам осуществляется по графикам, синхронизированным с производственным планом, что снижает простои склада на 18–22% [11] и повышает точность прогнозирования спроса [12].

Таблица 2

Прогноз ежемесячного спроса на продукцию завода «АвтоВАЗ»

Месяц	Прогноз спроса (шт.)	Комментарии / факторы влияния
Январь	25 444	Постновогодний период, низкая покупательная активность
Февраль	28 000	Небольшое оживление рынка, подготовка к весне
Март	34 349	Рост активности, старт сезонных распродаж
Апрель	38 167	Весеннее повышение спроса, первые отпуска
Май	40 711	Майские праздники, туристический сезон стимулирует продажи
Июнь	39 439	Летний период, часть покупателей в отпусках
Июль	44 527	Постепенный рост спроса (поездки на дальние расстояния, дачи и т. д.)
Август	49 605	Пик летнего сезона, подготовка к осени и обновлению автопарка
Сентябрь	53 434	Осенний рост продаж, деловой сезон, новые акции дилеров
Октябрь	50 889	Умеренный уровень спроса, стабилизация после сентябрьского пика
Ноябрь	48 255	Подготовка к зиме, часть клиентов обновляет авто перед холодами
Декабрь	47 180	Сезон скидок и новогодних акций, некоторые ждут выхода новых моделей
Итого	500 000	

Следует сформировать матрицу приоритетов по моделям [13]:

1. Lada Granta Седан – приоритет № 1, массовая модель.
2. Lada Vesta – приоритет № 2, перспективная средняя ценовая категория.
3. Lada Largus – приоритет № 3, семейный и коммерческий универсал.
4. Lada Niva Travel – приоритет № 4, внедорожная линейка.
5. Lada Aura – приоритет № 5, новая или узкоспециализированная модель.

Чтобы уложиться в 500 000 шт., следует удовлетворить спрос по самым приоритетным моделям и уменьшить выпуск менее приоритетных:

- ◆ Lada Granta Седан: спрос 200 000 – производим 200 000 (полностью удовлетворяет критериям).
- ◆ Lada Vesta: спрос 90 000 – производим 90 000 (полностью удовлетворяет).
- ◆ Lada Largus: спрос 100 000 – производим 100 000 (полностью удовлетворяет).
- ◆ Осталось «места» в плане: $500\,000 - (200\,000 + 90\,000 + 100\,000) = 110\,000$.
- ◆ Lada Niva Travel: спрос 60 000 – производим 60 000 (полностью удовлетворяем).
- ◆ Осталось «места»: $110\,000 - 60\,000 = 50\,000$.
- ◆ Lada Aura: спрос 70 000 – возможно произвести только 50 000.

Таким образом, итоговая таблица (табл. 3):

Таблица 3

Итоговая таблица после балансировки

Модель	Спрос (шт.)	План производства (шт.)	Разница
Lada Granta Седан	200 000	200 000	0
Lada Vesta	90 000	90 000	0
Lada Largus	100 000	100 000	0
Lada Niva Travel	60 000	60 000	0
Lada Aura	70 000	50 000	-20 000
Итого	520 000	500 000	-20 000

Все «лишние» 20 000 автомобилей «срезали» именно у Lada Aura как наименее приоритетной.

Для наглядности сводим плановые объёмы по каждой модели в таблицу (табл. 4):

Таблица 4

Плановые объёмы производства каждой модели

Модель	План производства (шт./год)
Lada Granta Седан	200 000
Lada Largus	100 000
Lada Aura	70 000
Lada Vesta	90 000
Lada Niva Travel	40 000
Итого	500 000

1. Lada Granta Седан. Самая массовая и популярная модель, ориентированная на широкий рынок и стабильный спрос.

2. Lada Largus. Семейный универсал, используемый также для коммерческих перевозок.

3. Lada Aura. Условная новая модель, создающая нишу между экономом и средним сегментом (цифры приведены в качестве примера).

4. Lada Vesta. Перспективная модель среднего класса, с относительно высоким спросом на внутреннем рынке.

5. Lada Niva Travel. Внедорожная модификация для любителей активного отдыха и эксплуатации на сложных дорогах.

Суммарный план годового производства — 500 000 автомобилей. Фактические объёмы выпуска внутри квартала или ежемесячно могут корректироваться в зависимости от спроса и доступности комплектующих.

План расхода материалов. Ниже показан упрощённый пример расчёта основных материалов и узлов на общее производство 500 000 автомобилей. В реальном планировании для каждой модели подбираются конкретные варианты двигателей, типов кузова, колёс, опций и т.д. (что приводит к более детальным расчётам).

Двигатели. Предположим, что каждая из перечисленных моделей в базовой версии комплектуется 1,6-литровым двигателем (код: ENG-01-02). Тогда при общем выпуске в 500 000 автомобилей:

- ♦ Всего двигателей: 500 000 шт.

При наличии модификаций (например, 1,8 л или 1,7 л для Niva Travel) объёмы по каждой версии двигателей распределялись бы пропорционально плану соответствующих моделей.

Кузовные детали. Для упрощения предположим, что завод штампует и окрашивает кузова самостоятельно для всего модельного ряда, а каждая машина требует 1 комплект кузовных деталей. Следовательно:

- ♦ Всего комплектов кузовных деталей: 500 000 комплектов.
- ♦ В учётных документах могут быть разделены на типы (Granta, Largus, Aura, Vesta, Niva), если технологически требуются разные штампы.

Шины и колёсные диски. Если каждая модель выходит с 16-дюймовыми шинами (4 шт. на автомобиль) плюс запаска (1 шт.) – всего 5 шт. на машину. В зависимости от политики предприятия запасное колесо может отличаться или отсутствовать (ремкомплект). Предположим, что у всех авто есть полноценная «запаска», тогда:

- ♦ Всего шин: $500\,000 \times 5 = 2\,500\,000$ шт.
- ♦ Аналогичное число дисков (если каждая шина идёт в сборе с диском).

Покрасочные работы. Учитывая, что каждый кузов обязательно проходит этап окраски (код: RB-01-07) и в общем объёме планируется 500 000 машин:

- ♦ Всего «работ по покраске кузова»: 500 000 усл. ед.

В учёт (если часть операций отдана подрядчикам) это может отражаться как закупка «работ по покраске» в количестве, соответствующем числу кузовов, или как внутренняя операция на собственных мощностях.

Предпродажная подготовка. Предусмотрим, что каждая машина перед отгрузкой проходит обязательную предпродажную подготовку – код: SV-01-01. Тогда:

- ♦ Всего услуг предпродажной подготовки: 500 000 усл. ед.

Поскольку «услуга» не хранится на складе, она списывается в затраты в момент её оказания. При этом важно обеспечить плановую загрузку постов ППП, достаточное количество работников и материалов (технических жидкостей, диагностического оборудования и т.д.).

Прочие материалы. В реальности перечень материалов и узлов гораздо шире:

- ♦ Силовая трансмиссия (коробки передач разных типов).
- ♦ Система выпуска отработанных газов, каталитические нейтрализаторы.
- ♦ Электроника (аккумуляторы, блоки ЭБУ, датчики и т. п.).
- ♦ Система кондиционирования и отопления (если предусмотрена).
- ♦ Внутренние детали салона (сиденья, обивка, рулевое колесо, панель приборов).
- ♦ Расходные материалы (краски, лаки, герметики, смазки, прокладки и пр.).

Каждая позиция рассчитывается по формуле:

$$Q_{\text{общ}} = N_{\text{план}} * q_{\text{норм}} , \quad (1)$$

$Q_{\text{общ}}$ – итоговое необходимое количество материалов;

$N_{\text{план}}$ – плановое количество выпускаемых автомобилей;

$q_{\text{норм}}$ – норма расхода материала на единицу продукции (на один автомобиль).

Этапы производства, производственные подразделения и рабочие центры. Для выполнения следующего задания создадим еще две номенклатурные позиции для описания этапов производства.

Номенклатура 1: «Кузовные детали Vesta» (тип: Товар)

Этапы производства

1. Штамповка – вырезка листовых заготовок и придание им формы (крылья, двери, капот, крышка багажника).
2. Контроль качества деталей – визуальный осмотр, измерения толщины металла, проверка правильности геометрии.
3. Промежуточная складировка – сортировка по местам хранения для дальнейшей передачи в сварочный или сборочный цех.

Производственные подразделения

1. Штамповочный цех. Основные операции: холодная штамповка, обрезка кромок.
2. Отдел технического контроля (ОТК). Проверка качества штампованных кузовных деталей.
3. Склад полуфабрикатов. Временное хранение заготовок перед дальнейшей обработкой или отгрузкой в сборочный цех.

Виды рабочих центров и рабочие центры

Тип: Прессовое оборудование

- ♦ *Рабочий центр 1:* Пресс-линия № 1 (крупная штамповка боковин кузова).
- ♦ *Рабочий центр 2:* Пресс-линия № 2 (штамповка дверей, капотов, багажников).

Тип: Линия окончательной обрезки

- ♦ *Рабочий центр 3:* Обрезной участок для точной калибровки кромок деталей.

Тип: Контрольный пост

- ♦ *Рабочий центр 4:* Станция визуального и инструментального контроля, где сотрудники ОТК проводят замеры, проверяют соответствие детализированным чертежам.

Номенклатура 2: «Двигатель 1,6 л / ENG-01-02» (тип: Товар)

Этапы производства

1. Механическая обработка блоков: фрезеровка, расточка цилиндров, нарезка резьбы на деталях.
2. Сборка двигателей: установка коленчатого вала, поршней, ГБЦ, навесных агрегатов.
3. Обкатка и тестирование: прогон на тестовых стендах, проверка на герметичность, мощность, отсутствие вибраций и шумов.

Производственные подразделения

1. Механический цех. Проводит первичную и точную обработку блоков цилиндров, ГБЦ и других деталей.
2. Сборочный участок двигателей. Отдельная зона, где блоки комплектуются в готовые моторы (включая навесные узлы).
3. Испытательный отдел. Специализированные стенды для проверки, обкатки и настройки каждого двигателя.

Виды рабочих центров и рабочие центры

Тип: Механические обрабатывающие центры

- ♦ *Рабочий центр 1:* Токарно-фрезерная линия (расточка блоков, чистовая обработка).
- ♦ *Рабочий центр 2:* Линия сверлильно-резьбовых станков (обработка отверстий, нарезка резьбы).

Тип: Сборочная линия двигателей

- ♦ *Рабочий центр 3:* Роботизированная сборочная линия (автоматизированная укладка коленвала, поршневой группы).
- ♦ *Рабочий центр 4:* Ручной участок доводки (установка ГБЦ, выпускного/впускного коллектора).

Тип: Испытательный стенд

- ♦ *Рабочий центр 5:* Обкаточная линия (проверка работоспособности каждого двигателя при различных нагрузках).

Номенклатура 3: «Lada Vesta» (тип: Товар, готовый автомобиль)

Этапы производства

1. Сварка кузова: соединение штампованных деталей (пол, крыша, боковины, двери и т.д.) в каркас кузова.
2. Окраска кузова: грунтовка, нанесение слоя краски и лака, сушка и контроль качества окрашенного кузова.
3. Основная сборка: установка двигателя, ходовой части, трансмиссии, салона, электрики, колёсных узлов.
4. Тестирование и предпродажная подготовка: проверка на специальных стендах, устранение дефектов, оформление документов, финальный осмотр.

Производственные подразделения

1. Сварочный цех. Формирование кузова из отдельных штампованных панелей.
2. Покрасочный цех. Полный цикл подготовки поверхности и нанесения лакокрасочного покрытия.
3. Главный сборочный цех. Агрегатная сборка автомобиля: двигатель, подвеска, салон, электрика и т.д.
4. Отдел технического контроля. Выполнение ходовых испытаний, электрических и диагностических проверок.
5. Отдел предпродажной подготовки. Финальная проверка перед передачей на склад готовой продукции.

Виды рабочих центров и рабочие центры

Тип: Роботизированные сварочные линии

- ♦ *Рабочий центр 1:* Линия сварки каркаса кузова (роботы-манипуляторы).
- ♦ *Рабочий центр 2:* Ручной пост сварки сложных узлов (например, стыков дверей).

Тип: Покрасочный комплекс

- ♦ *Рабочий центр 3:* Камера грунтовки и катафорезной обработки.
- ♦ *Рабочий центр 4:* Камера финишной покраски (нанесение базовой краски и лака).

Тип: Сборочные линии

- ♦ *Рабочий центр 5:* Линия установки силового агрегата (двигатель + трансмиссия).
- ♦ *Рабочий центр 6:* Линия установки салона и электрооборудования (панель приборов, кресла, жгуты проводки).

Тип: Тестовые и диагностические стенды

- ♦ *Рабочий центр 7:* Ходовой стенд (проверка динамических характеристик, рулевого управления).
- ♦ *Рабочий центр 8:* Электронная диагностика (чтение кодов ошибок, проверка бортовых систем).

Тип: Участок предпродажной подготовки

- ♦ *Рабочий центр 9:* Полировка кузова, заправка техническими жидкостями, контроль основных функций.

Проектирование ресурсных спецификаций и автовыбор материала

Для начала следует создать ресурсную спецификацию «Lada Vesta».

Укажем этапы во вкладке «Производственный процесс».

Для расчета требуется создать 12 характеристик номенклатуры.

Рассчитывать количество комплектов кузовных деталей, комплектов шин и двигателей будем исходя из количества автомобилей по формуле:

$$Q_{\text{компл}} = N_{\text{авто}} \quad (2)$$

где $Q_{\text{компл}}$ – необходимое количество комплектов кузовных деталей, комплектов шин или двигателей; $N_{\text{авто}}$ – количество выпускаемых автомобилей.

Рассчитывать количество лакокрасочных материалов можно по формуле:

$$Q_{\text{ЛКМ}} = 0,5 * V_{\text{кузова}} \quad (3)$$

где $Q_{\text{ЛКМ}}$ – необходимое количество лакокрасочных материалов (литров); $V_{\text{кузова}}$ – объём кузова автомобиля (м^3); $0,5$ – числовой коэффициент, отвечающий за норматив расхода лакокрасочных материалов на 1 м^3 .

То есть посчитанный объем кузова умножаем на требуемый литраж на покраску одного кубического метра.

Моторное масло рассчитывается по формуле:

$$Q_{\text{масла}} = 4 * N_{\text{авто}}, \quad (4)$$

где $Q_{\text{масла}}$ – необходимое количество моторного масла (литров); $N_{\text{авто}}$ – количество выпускаемых автомобилей; 4 – числовой коэффициент нормативного расхода моторного масла на один автомобиль (литры).

Сценарии годового планирования. Для выполнения этого задания для начала создадим товарные категории по типу кузова и укажем их для каждой номенклатуры.

Проведение расчета сумм продаж исходя из нормативов распределения. Для выполнения данной задачи в Excel требуется рассчитать сумму продаж исходя из заданных нормативов распределения и трех параметров.

Реализация документооборота продаж при оформлении операций оптовой продажи: коммерческое предложение клиенту, заказ, реализация, счет.

Обсуждение. Результаты подтверждают, что КИС обеспечивает прозрачность распределения ресурсов [14] и снижает операционные издержки [15]. Важность внедрения таких систем в современной промышленности обоснована необходимостью повышения конкурентоспособности, автоматизации управления ресурсами и оптимизации производственных процессов. Полученные результаты демонстрируют существенное улучшение операционной эффективности: в частности, было достигнуто сокращение времени обработки заказов на 30%, а также реализована балансировка производственного плана с учетом приоритетных моделей автомобилей, таких как Lada Granta. Однако ключевым вызовом остается обучение сотрудников работе с ERP-интерфейсами [16]. Отмечена важность гибридных облачных решений для масштабируемости системы [17].

Необходимо подчеркнуть, что внедрение КИС на практике сопровождалось рядом вызовов и ограничений. К ним относятся традиционные проблемы предприятий, такие как несогласованность подразделений и устаревшие методы планирования, требующие значительных организационных и методологических изменений. Особое внимание было уделено важности профессионального обучения сотрудников и использованию гибридных облачных решений, которые способствуют успешной интеграции новых технологий и систем в существующую инфраструктуру.

Представленное исследование подтверждает значительный потенциал корпоративных информационных систем для трансформации производственных и управленческих процессов на крупных промышленных предприятиях. Вместе с тем, дальнейшие исследования могут быть направлены на углубленное изучение специфических аспектов взаимодействия ERP-систем с другими корпоративными приложениями, а также на оценку долгосрочных эффектов от внедрения данных технологий на промышленном предприятии.

Заключение. В рамках данного исследования была успешно реализована разработка и внедрение элементов корпоративной информационной системы на промышленном предприятии завода «АвтоВАЗ» [18]. Внедрение системы «1С:ERP» продемонстрировало очевидные преимущества в области управления ресурсами и оптимизации производственного планирования, включая снижение времени обработки заказов и повышение прозрачности распределения ресурсов.

Полученные результаты подтвердили высокую эффективность предложенных решений и их значительный потенциал для повышения конкурентоспособности предприятия в условиях современного рынка. Также были определены ключевые факторы успешного внедрения КИС, такие как обучение сотрудников, внедрение современных технологий и применение гибридных облачных решений [19].

Таким образом, разработанные методы и подходы могут быть успешно адаптированы и применены на других предприятиях автомобильной и машиностроительной отрасли, способствуя улучшению их производственных и управленческих процессов [20].

Выводы. В рамках данного исследования была успешно реализована разработка и внедрение элементов корпоративной информационной системы на промышленном предприятии завода «АвтоВАЗ». Внедрение системы «1С:ERP» продемонстрировало очевидные преимущества в области управления ресурсами и оптимизации производственного планирования, включая снижение времени обработки заказов и повышение прозрачности распределения ресурсов.

Полученные результаты подтвердили высокую эффективность предложенных решений и их значительный потенциал для повышения конкурентоспособности предприятия в условиях современного рынка. Также были определены ключевые факторы успешного внедрения КИС, такие как обучение сотрудников, внедрение современных технологий и применение гибридных облачных решений.

Таким образом, разработанные методы и подходы могут быть успешно адаптированы и применены на других предприятиях автомобильной и машиностроительной отрасли, способствуя улучшению их производственных и управленческих процессов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Шитова Татьяна Федоровна, Молодецкая Светлана Федоровна.* 1С:ERP – эффективный инструмент развития цифровой экономики // Муниципалитет: экономика и управление. – 2021. – № 3 (36). – URL: <https://cyberleninka.ru/article/n/1s-erp-effektivnyy-instrument-razvitiya-tsifrovoy-ekonomiki> (дата обращения: 25.12.2024).
2. *Альбекова З.М., Балабина А.С., Коротченко В.П.* Современные корпоративные информационные системы и корпоративные информационные порталы // Инженерный вестник Дона. – 2018. – № 1 (48). – URL: <https://cyberleninka.ru/article/n/sovremennye-korporativnye-informatsionnye-sistemy-i-korporativnye-informatsionnye-portaly> (дата обращения: 25.12.2024).
3. ВИ-технологии и корпоративные информационные системы в оптимизации бизнес-процессов цифровой экономики // Матер. XI Международной научно-практической очно-заочной конференции, Екатеринбург, 01 декабря 2023 года. – Екатеринбург: Уральский государственный экономический университет, 2024. – 199 с. – EDN BBBICA.
4. *Ковалев В.Е., Новикова К.В., Добровлянин В.Д.* Внедрение ERP-систем малыми и средними компаниями в России: барьеры и перспективы // Управленец. – 2023. – Т. 14, № 6. – С. 77-90. – DOI 10.29141/2218-5003-2023-14-6-6. – EDN EPLGEJ.
5. *Галкин А.В., Смолин А.В., Неволлина Е.М.* Управление производственным риском как элемент проектирования системы обеспечения безопасности труда горнодобывающего предприятия, обеспечивающий надежность ее функционирования // Горная промышленность. – 2022. – № S1. – С. 86-94. – DOI 10.30686/1609-9192-2022-S-86-94. – EDN KEQKIQ.
6. *Шайдуллин А.И., Жусупов Н.Б.* Особенности оценки инвестиционной привлекательности инновационных рынков в условиях цифровой трансформации экономики // Информационное общество. – 2025. – № 1. – С. 70-78. – DOI 10.52605/16059921_2025_01_70. – EDN CCSQTT.
7. *Лукашов Н.В.* Актуальные подходы к ресурсному управлению рисками инновационных проектов // Вестник Санкт-Петербургского университета. Экономика. – 2023. – Т. 39, № 2. – С. 217-247. – DOI 10.21638/spbu05.2023.204. – EDN BEYDRK.
8. *Кочетков Е.А.* Проектное управление как инструмент внедрения изменений в организациях // Инновационные технологии управления: Сб. статей по материалам VIII Всероссийской научно-практической конференции, Нижний Новгород, 17–18 ноября 2021 года. – Нижний Новгород: Федеральное государственное бюджетное образовательное учреждение высшего образования "Нижегородский государственный педагогический университет имени Козьмы Минина", 2021. – С. 49-51. – EDN HOLVUO.
9. *Остроумов А.П.* Реализация принципов практико-ориентированного подхода к выполнению выпускной квалификационной работы бакалавра по направлению 38.03.05 Бизнес информатика на основе сервиса "1С:Предприятие 8" для учебных заведений через интернет" // Новые информационные технологии в образовании: Инновации в экономике и образовании на базе технологических решений 1С: Сб. научных трудов 17-й Международной научно-практической конференции, Москва, 31 января – 01 2017 года. Том Часть 1. – М.: Общество с ограниченной ответственностью "1С-Пабблишинг", 2017. – С. 410-413. – EDN YQBYZS.
10. *Баранова С.В., Коломейченко А.С., Баранов Ю.Н.* Оптимизация деятельности региональных автомобильных дилерских центров // Мир транспорта и технологических машин. – 2018. – № 4 (63). – С. 131-139. – EDN VQLKEC.

11. Михелаишвили А.М., Трейман М.Г. Корпоративные информационные системы как элемент управления процессами // Молодежь, образование и наука XXI века: Матер. научно-практической конференции студентов и аспирантов, посвящённой памяти заслуженного деятеля науки РФ, профессора В.С. Соминского, Санкт-Петербург, 18 апреля 2024 года. – СПб.: Санкт-Петербургский государственный университет промышленных технологий и дизайна, 2024. – С. 120-122. – EDN EWXSKD.
12. Шипиловских Н.А. Как влияет внедрение корпоративных информационных систем на экономику // Современные вопросы естествознания и экономики: Сб. трудов VI Международной научно-практической конференции, Прокопьевск, 21 марта 2024 года. – Прокопьевск: Кузбасский государственный технический университет им. Т.Ф. Горбачева, 2024. – С. 192-194. – EDN MMUMYL.
13. Лямцева А.А. Проблема импортозамещения ERP-систем на российском рынке // Производственный менеджмент: теория, методология, практика. – 2016. – № 8. – С. 49-54. – EDN XEGSTJ.
14. Десятирикова Е.Н., Романов А.С., Воробьева И.Ю. Анализ применимости отечественных ERP продуктов // Системный анализ в проектировании и управлении: Сб. научных трудов XIV Международной научно-практической конференции, Санкт-Петербург, 29 июня – 01 июля 2010 года. Т. 2. – СПб.: Центральный экономико-математический институт РАН, 2010. – С. 111-113. – EDN VNBKMD.
15. Маслов Е.А. Корпоративная информационная Система в системе корпоративного управления // Теория и практика современной науки. – 2024. – № 5 (107). – С. 108-110. – EDN DCHXBR.
16. Кузина Д.А. Бизнес-аналитика и повышение эффективности внедрения ERP-системы организации // Развитие региональной инновационной системы на основе импортоопережения и заделов интеллектуальной собственности: Сб. материалов Международной научно-практической конференции, Самара, 21 ноября 2024 года. – Самара: ООО "Самарама", 2024. – С. 215-233. – EDN VWYZGA.
17. Бельмас С.М. Разработка и внедрение ERP-систем // Шумпетеровские чтения. – 2018. – Т. 1. – С. 48-60. – EDN ZCTJJR.
18. Копань А.О., Великанова Л.О. Информатизация производственного планирования: типовые модули производственного планирования // Цифровизация экономики: направления, методы, инструменты: Сб. материалов II всероссийской научно-практической конференции, Краснодар, 20–24 января 2020 года. – Краснодар: Кубанский государственный аграрный университет имени И.Т. Трубилина, 2020. – С. 281-283. – EDN GMFIYM.
19. Доронкин В.Г., Ельцов В.В. ВАЗ и ВУЗ. Исторические параллели. Опыт реализации стратегии развития 2020 // Инженерное образование. – 2016. – № 19. – С. 116-121. – EDN WWULZT.
20. Дадян Э.Г. Конфигурирование и моделирование в системе "1С: Предприятие": учебник. – М.: Издательский дом "Вузовский учебник", 2018. – 417 с. – (Высшее образование: Магистратура (ФУ)). – ISBN 978-5-9558-0581-8. – DOI 10.12737/textbook_5a0c1bcccc76f5.69529307. – EDN YMGIJS.

REFERENCES

1. Shitova Tat'yana Fedorovna, Molodetskaya Svetlana Fedorovna. 1S:ERP – effektivnyy instrument razvitiya tsifrovoy ekonomiki [1C:ERP as an Effective Tool for Digital Economy Development], *Munitsipalitet: ekonomika i upravlenie* [Municipality: Economics and Management], 2021, No. 3 (36). Available at: <https://cyberleninka.ru/article/n/1s-erp-effektivnyy-instrument-razvitiya-tsifrovoy-ekonomiki> (accessed 25 December 2024).
2. Al'bekova Z.M., Balabina A.S., Korotchenko V.P. Sovremennyye korporativnyye informatsionnyye sistemy i korporativnyye informatsionnyye portaly [Modern corporate information systems and corporate information portals], *Inzhenernyy vestnik Dona* [Engineering journal of Don], 2018, No. 1 (48). Available at: <https://cyberleninka.ru/article/n/sovremennyye-korporativnyye-informatsionnyye-sistemy-i-korporativnyye-informatsionnyye-portaly> (accessed 25 December 2024).
3. BI-tehnologii i korporativnyye informatsionnyye sistemy v optimizatsii biznes-protsessov tsifrovoy ekonomiki [BI technologies and corporate information systems in optimizing digital economy business processes], *Mater. XI Mezhdunarodnoy nauchno-prakticheskoy ochno-zaochnoy konferentsii, Ekaterinburg, 01 dekabrya 2023 goda* [Proceedings of the XI International Scientific-Practical Conference (Hybrid Format), Yekaterinburg, December 1, 2023]. Ekaterinburg: Ural'skiy gosudarstvennyy ekonomicheskii universitet, 2024, 199 p. EDN BBBICA.
4. Kovalev V.E., Novikova K.V., Dobrovlyanin V.D. Vnedrenie ERP-sistem malymi i srednimi kompaniyami v Rossii: bar'ery i perspektivy [Implementation of ERP systems by small and medium-sized companies in Russia: barriers and prospects], *Upravlenets* [Manager], 2023, Vol. 14, No. 6, pp. 77-90. DOI 10.29141/2218-5003-2023-14-6-6. – EDN EPLGEJ.

5. Galkin A.V., Smolin A.V., Nevolina E.M. Upravlenie proizvodstvennym riskom kak element proektirovaniya sistemy obespecheniya bezopasnosti truda gornodobyvayushchego predpriyatiya, obespechivayushchiy nadezhnost' ee funktsionirovaniya [Production risk management as an element of occupational safety system design for mining enterprises], *Gornaya promyshlennost'* [Mining Industry], 2022, No. S1, pp. 86-94. DOI 10.30686/1609-9192-2022-1S-86-94. EDN KEQKIQ.
6. Shaydullin A.I., Zhusupov N.B. Osobennosti otsenki investitsionnoy privlekatel'nosti innovatsionnykh rynkov v usloviyakh tsifrovoy transformatsii ekonomiki [Features of assessing investment attractiveness of innovation markets in the context of digital economic transformation], *Informatsionnoe obshchestvo* [Information Society], 2025, No. 1, pp. 70-78. DOI 10.52605/16059921_2025_01_70. – EDN CCSQTT.
7. Lukashov N.V. Aktual'nye podkhody k resursnomu upravleniyu riskami innovatsionnykh projektov [Current approaches to resource risk management in innovation projects], *Vestnik Sankt-Peterburgskogo universiteta. Ekonomika* [Bulletin of Saint Petersburg University. Economics], 2023, Vol. 39, No. 2, pp. 217-247. DOI 10.21638/spbu05.2023.204. EDN BEYDRK.
8. Kochetkov E.A. Proektnoe upravlenie kak instrument vnedreniya izmeneniy v organizatsiyakh [Project management as a tool for implementing organizational change], *Innovatsionnye tekhnologii upravleniya: Sb. statey po materialam VIII Vserossiyskoy nauchno-prakticheskoy konferentsii, Nizhniy Novgorod, 17–18 noyabrya 2021 goda* [Innovative Management Technologies: Proceedings of the VIII All-Russian Scientific-Practical Conference, Nizhny Novgorod, November 17–18, 2021]. Nizhniy Novgorod: Federal'noe gosudarstvennoe byudzhethoe obrazovatel'noe uchrezhdenie vysshego obrazovaniya "Nizhegorodskiy gosudarstvennyy pedagogicheskiy universitet imeni Koz'my Minina", 2021, pp. 49-51. EDN HOLVUO.
9. Ostroumov A.P. Realizatsiya printsipov praktiko-orientirovannogo podkhoda k vypolneniyu vypusknoy kvalifikatsionnoy raboty bakalavra po napravleniyu 38.03.05 Biznes informatika na osnove servisa "1S:Predpriyatie 8" dlya uchebnykh zavedeniy cherez internet" [Implementation of Practice-Oriented Principles in Bachelor's Theses for Business Informatics Using the 1C:Enterprise 8 Educational Service], *Novye informatsionnye tekhnologii v obrazovanii: Innovatsii v ekonomike i obrazovanii na baze tekhnologicheskikh resheniy 1S: Sb. nauchnykh trudov 17-y Mezhdunarodnoy nauchno-prakticheskoy konferentsii, Moskva, 31 yanvarya – 01 2017 goda* [New Information Technologies in Education: Innovations in Economy and Education Based on 1C Solutions: Proceedings of the 17th International Scientific-Practical Conference, Moscow, January 31 – February 1, 2017]. Part 1. Moscow: Obshchestvo s ogranichennoy otvetstvennost'yu "1S-Publishing", 2017, pp. 410-413. EDN YQBYZS.
10. Baranova S.V., Kolomeychenko A.S., Baranov Yu.N. Optimizatsiya deyatel'nosti regional'nykh avtomobil'nykh dilerskikh tsentrov [Optimization of Regional Automotive Dealership Operations], *Mir transporta i tekhnologicheskikh mashin* [World of Transport and Technological Machines], 2018, No. 4 (63), pp. 131-139. EDN VQLKEC.
11. Mikhelashvili A.M., Treyman M.G. Korporativnye informatsionnye sistemy kak element upravleniya protsessami [Corporate information systems as an element of process management], *Molodezh', obrazovanie i nauka XXI veka: Mater. nauchno-prakticheskoy konferentsii studentov i aspirantov, posvyashchennoy pamyati zasluzhennogo deyatelya nauki RF, professora V.S. Sominskogo, Sankt-Peterburg, 18 aprelya 2024 goda* [Youth, Education, and Science of the 21st Century: Proceedings of the Conference Dedicated to the Memory of Professor V.S. Sominsky, Saint Petersburg, April 18, 2024]. Saint Petersburg: Sankt-Peterburgskiy gosudarstvennyy universitet promyshlennykh tekhnologiy i dizayna, 2024, pp. 120-122. EDN EWXSKD.
12. Shipilovskikh N.A. Kak vliyaet vnedrenie korporativnykh informatsionnykh sistem na ekonomiku [The impact of corporate information systems implementation on the economy], *Sovremennye voprosy estestvoznaniya i ekonomiki: Sb. trudov VI Mezhdunarodnoy nauchno-prakticheskoy konferentsii, Prokop'evsk, 21 marta 2024 goda* [Modern Issues of Natural Sciences and Economics: Proceedings of the VI International Scientific-Practical Conference, Prokopyevsk, March 21, 2024]. Prokop'evsk: Kuzbasskiy gosudarstvennyy tekhnicheskiiy universitet im. T.F. Gorbacheva, 2024, pp. 192-194. EDN MMUMYL.
13. Lyamtseva A.A. Problema importozameshcheniya ERP-sistem na rossiyskom rynke [The problem of import substitution of ERP systems in the russian market], *Proizvodstvennyy menedzhment: teoriya, metodologiya, praktika* [Production Management: Theory, Methodology, Practice], 2016, No. 8, pp. 49-54. EDN XEGSTJ.
14. Desyatirikova E.N., Romanov A.S., Vorob'eva I.Yu. Analiz primenimosti otechestvennykh ERP produktov [Applicability analysis of domestic ERP products], *Sistemnyy analiz v proektirovanii i upravlenii: Sb. nauchnykh trudov XIV Mezhdunarodnoy nauchno-prakticheskoy konferentsii, Sankt-Peterburg, 29 iyunya – 01 2010 goda* [System Analysis in Design and Management: Proceedings of the XIV International Scientific-Practical Conference, Saint Petersburg, June 29 – July 1, 2010]. Vol. 2. Saint Petersburg: TSentral'nyy ekonomiko-matematicheskiiy institut RAN, 2010, pp. 111-113. EDN VNBKMD.

15. Maslov E.A. Korporativnaya informatsionnaya Sistema v sisteme korporativnogo upravleniya [Corporate information system in corporate governance], *Teoriya i praktika sovremennoy nauki* [Theory and Practice of Modern Science], 2024, No. 5 (107), pp. 108-110. EDN DCHXBR.
16. Kuzina D.A. Biznes-analitika i povyshenie effektivnosti vnedreniya ERP-sistemy organizatsii [Business analytics and improving the efficiency of ERP system implementation], *Razvitie regional'noy innovatsionnoy sistemy na osnove importooperezheniya i zadelov intellektual'noy sobstvennosti: Sb. materialov Mezhdunarodnoy nauchno-prakticheskoy konferentsii, Samara, 21 noyabrya 2024 goda* [Development of Regional Innovation Systems Based on Import Substitution and Intellectual Property: Proceedings of the International Scientific-Practical Conference, Samara, November 21, 2024]. Samara: OOO "Samarama", 2024, pp. 215-233. EDN VWYZGA.
17. Bel'mas S.M. Razrabotka i vnedrenie ERP-sistem [Development and Implementation of ERP Systems], *Shumpeterovskie chteniya* [Schumpeterian Readings.], 2018, Vol. 1, pp. 48-60. EDN ZCTJJR.
18. Kopan' A.O., Velikanova L.O. Informatizatsiya proizvodstvennogo planirovaniya: tipovye moduli proizvodstvennogo planirovaniya [Informatization of production planning: standard modules], *Tsifrovizatsiya ekonomiki: napravleniya, metody, instrumenty: Sb. materialov II vserossiyskoy nauchno-prakticheskoy konferentsii, Krasnodar, 20–24 yanvarya 2020 goda* [Digitalization of the Economy: Directions, Methods, Tools: Proceedings of the II All-Russian Scientific-Practical Conference, Krasnodar, January 20–24, 2020]. Krasnodar: Kubanskiy gosudarstvennyy agrarnyy universitet imeni I.T. Trubilina, 2020, pp. 281-283. EDN GMFIYM.
19. Doronkin V.G., El'tsov V.V. VAZ i VUZ. Istoricheskie paralleli. Opyt realizatsii strategii razvitiya 2020 [VAZ and higher education: historical parallels. Experience in implementing the 2020 development strategy], *Inzhenernoe obrazovanie* [Engineering Education], 2016, No. 19, pp. 116-121. EDN WWULZT.
20. Dadyan E.G. Konfigurirovanie i modelirovanie v sisteme "1S: Predpriyatie": uchebnik [Configuration and modeling in the 1c:enterprise system: textbook]. Moscow: Izdatel'skiy dom "Vuzovskiy uchebnik", 2018. – 417 s. – (Vysshee obrazovanie: Magistratura (FU)). ISBN 978-5-9558-0581-8. DOI 10.12737/textbook_5a0c1bcccc76f5.69529307. EDN YMGIJS.

Зорькин Дмитрий Юрьевич – Волгоградский государственный технический университет; e-mail: mosh285@gmail.com; г. Волгоград, Россия; преподаватель кафедры ПМ.

Богнюков Александр Алексеевич – Волгоградский государственный технический университет; e-mail: bogniukovv@gmail.com; г. Волгоград, Россия; студент.

Кожанова Татьяна Евгеньевна – Волгоградский государственный технический университет; e-mail: mitina.te@yandex.ru; г. Волгоград, Россия; старший преподаватель кафедры ПМ.

Zorkin Dmitry Yuryevich – Volgograd State Technical University; e-mail: mosh285@gmail.com; Volgograd, Russia; teacher of PM Department.

Bognyukov Alexander Alekseevich – Volgograd State Technical University; e-mail: bogniukovv@gmail.com; Volgograd, Russia; student.

Kozhanova Tatiana Evgenievna – Volgograd State Technical University; e-mail: mitina.te@yandex.ru; Volgograd, Russia; senior lecturer at the Department of PM.

УДК 621.373+519.622

DOI 10.18522/2311-3103-2026-1-112-122

А.М. Пилипенко

ПРИМЕНЕНИЕ ГИБРИДНЫХ МЕТОДОВ ЧИСЛЕННОГО РЕШЕНИЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ ДЛЯ АНАЛИЗА АВТОКОЛЕБАТЕЛЬНЫХ ЦЕПЕЙ С РАЗЛИЧНОЙ ДИНАМИКОЙ

Обеспечение точности и устойчивости компьютерного моделирования радиоэлектронных устройств является важной задачей при их проектировании. Наибольшие трудности при компьютерном моделировании радиоэлектронных устройств возникают в случае анализа автоколебательных цепей, поскольку математические модели таких цепей могут быть жесткими и осциллирующими одновременно. Целью данной работы является обоснование эффективного численного метода решения обыкновенных дифференциальных уравнений, имеющего более высокую точность

анализа во временной области различных видов автоколебательных цепей по сравнению с известными методами. Предлагаемый метод является гибридным методом и базируется на известных методах Гира и трапеций, использующихся в пакетах схемотехнического моделирования радиоэлектронных устройств. Для оценки точности предлагаемого метода и известных методов использовалась обобщенная модель автоколебательной цепи, для которой было определено аналитическое решение в установившемся режиме работы. Точность численного решения определялась на основе анализа погрешностей оценки основных параметров колебательного процесса – амплитуды и частоты колебаний. Сравнительный анализ погрешностей оценки амплитуды и частоты колебаний автогенераторов в различных режимах работы показывает высокую эффективность предлагаемого гибридного метода как для анализа генераторов гармонических колебаний, так и для анализа генераторов релаксационных колебаний. Дальнейшее повышение точности гибридного метода возможно за счет применения неявных методов Рунге-Кутты высокого порядка точности, а именно подклассов Радо IIA и Лобатто IIIA, которые обладают L- и P-устойчивостью соответственно. Следует отметить, что с ростом порядка точности неявных методов Рунге-Кутты возрастает вычислительная трудоемкость данных методов, но для подклассов Радо IIA и Лобатто IIIA рост вычислительной трудоемкости будет минимальным.

Численный метод; обыкновенное дифференциальное уравнение; автогенератор; гармоническое колебание; релаксационное колебание; погрешность моделирования.

A.M. Pilipenko

APPLICATION OF HYBRID METHODS FOR NUMERICAL SOLVING OF ORDINARY DIFFERENTIAL EQUATIONS FOR ANALYSIS OF SELF-OSCILLATING CIRCUITS WITH VARIOUS DYNAMICS

Ensuring the accuracy and stability of computer simulation of electronic devices is an important problem in their design. The greatest difficulties in simulation of electronic devices arise in the case of the analysis of self-oscillating circuits, since mathematical models of such circuits can be stiff and oscillating at the same time. The aim of this work is to develop an efficient numerical method for solving ordinary differential equations that provides higher accuracy of time domain analysis for various types of autogenerators compared to existing methods. The proposed method is a hybrid method and is based on the well-known Gear and trapezoidal methods used in simulators of electronic circuits. To evaluate the accuracy of the proposed method and known methods a generalized model of a self-oscillating circuit was used for which an analytical solution was determined in the steady-state operating mode. The accuracy of the numerical solution was determined based on the analysis of errors in estimating the main parameters of the oscillatory process – the amplitude and frequency of oscillations. A comparative analysis of errors in estimating the amplitude and frequency of oscillations in autogenerators demonstrates the high efficiency of the proposed hybrid method for analyzing both harmonic oscillators and relaxation oscillators. A further increase in the accuracy of the hybrid method is possible using implicit Runge-Kutta methods (Rado IIA and Lobatto IIIA subclasses), which have L- and P-stability, respectively. It should be noted that with an increase in the order of accuracy of implicit Runge-Kutta methods, the computational complexity of these methods increases, but for the Rado IIA and Lobatto IIIA subclasses the increase in computational complexity will be minimal.

Numerical method; ordinary differential equation; autogenerator; harmonic oscillation; relaxation oscillation; simulation error.

Введение. В современных пакетах схемотехнического моделирования (SPICE-симуляторах), в том числе отечественных, для автоматизированного анализа цепей во временной области используются методы Гира и метод трапеций [1, 2]. Указанные методы имеют минимальную вычислительную сложность среди неявных методов численного интегрирования обыкновенных дифференциальных уравнений (ОДУ). Методы Гира обладают L-устойчивостью и могут использоваться для анализа жестких систем, постоянные времени которых отличаются на порядок и более. Метод трапеций является P-устойчивым и обеспечивает максимальную точность анализа осциллирующих систем, процессы в которых близки к гармоническим колебаниям [3].

Несмотря на перечисленные выше достоинства методов Гира и трапеций в настоящее время продолжается разработка более эффективных методов численного решения ОДУ, в том числе применительно к задачам радиоэлектроники [4–6]. Действительно, из-

вестные методы имеют ряд недостатков, которые представлены, например, в работах [7–9]. Наибольшие трудности при использовании методов Гира и трапеций возникают при анализе автоколебательных цепей.

На рис. 1 и 2 представлены результаты численного анализа во временной области различных видов автогенераторов, полученные в работе [9]. Схемы анализируемых автогенераторов были описаны в работе [10] и адаптированы для симулятора Multisim в работе [9].

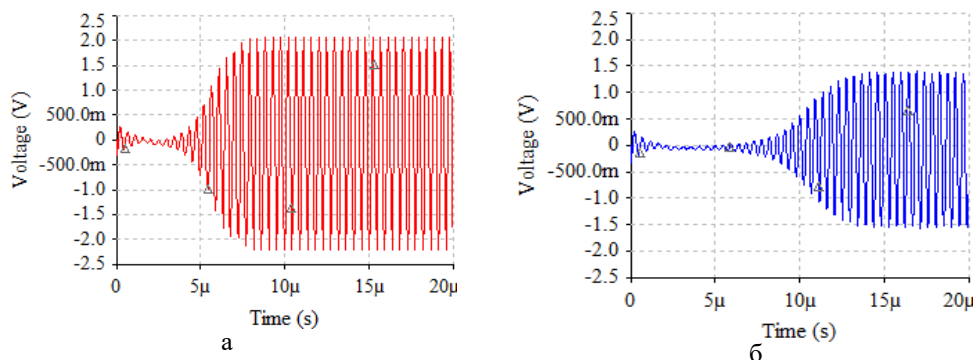


Рис. 1. Результаты численного анализа генератора гармонических колебаний, полученные методами трапеций (а) и Гира (б) [9]

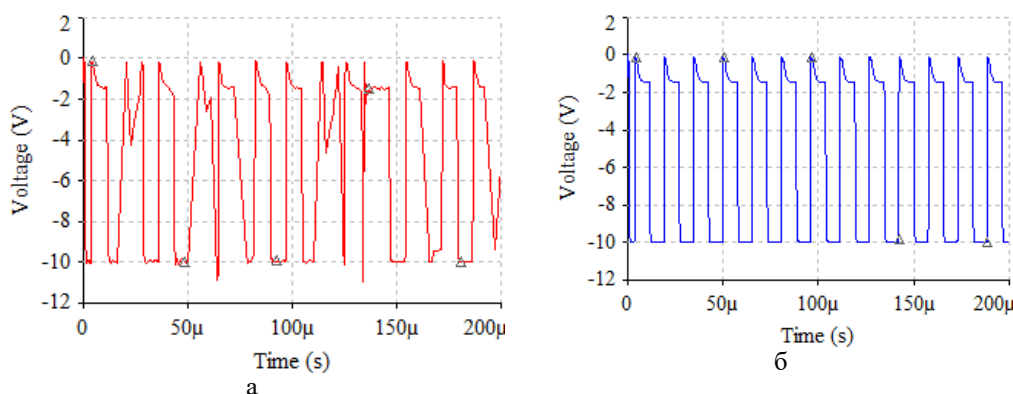


Рис. 2. Результаты численного анализа генератора релаксационных колебаний, полученные методами трапеций (а) и Гира (б) [9]

Из рис. 1 видно, что использование методов Гира для анализа генераторов гармонических колебаний приводит к получению колебания, у которого установившееся значение амплитуды отличается примерно в 2 раза от истинного значения. Следует отметить, что метод Гира также может приводить к затухающему численному решению при численном анализе генераторов гармонических колебаний. В свою очередь, как следует из рис. 2, использование метода трапеций для анализа генераторов релаксационных колебаний приводит к искажению решения и появлению паразитных колебаний.

Целью данной работы является обоснование гибридного метода численного интегрирования обыкновенных дифференциальных уравнений, обеспечивающего высокую точность и устойчивость моделирования во временной области автогенераторов гармонических и релаксационных колебаний.

В соответствии с поставленной целью в работе решаются следующие задачи:

- ♦ описание обобщенной модели автоколебательной цепи и методики тестирования численных методов;

- ♦ обоснование разностной схемы гибридного метода на основе методов Гира и трапеций;
- ♦ сравнительный анализ точности известных численных методов и гибридных методов.

1. Обобщенная модель автоколебательной цепи. Для количественной оценки точности численных методов тестовая задача должна иметь аналитическое решение. В данной работе в качестве тестовой задачи используется обобщенная модель автоколебательной цепи, предложенная в работе [7], для которой было определено точное решение в стационарном режиме, а также найдены установившиеся значения амплитуд и частот с пятнадцатью точными значащими цифрами для случаев гармонических и релаксационных колебаний [9]. Эквивалентная схема рассматриваемой автоколебательной цепи показана на рис. 3.

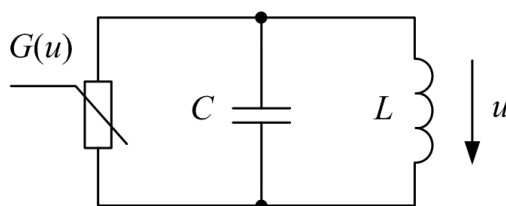


Рис. 3. Схемная модель автоколебательной цепи

В схемной модели на рис. 3 обозначены линейные элементы контура L и C , а также нелинейная проводимость, которая описывается следующим выражением [9]

$$G(u) = \begin{cases} G_1, & u < -U_0, u > U_0; \\ -G_0, & -U_0 \leq u \leq U_0, \end{cases} \quad (1)$$

где u – напряжение на выходе цепи; G_0 и G_1 – модули дифференциальной проводимости при $G(u) \leq 0$ и $G(u) > 0$; $\pm U_0$ – напряжения переключения $G(u)$.

Математическая модель автоколебательной цепи, показанной на рис. 3, представляет собой нелинейную систему ОДУ:

$$\begin{aligned} \frac{du}{dt} &= -\frac{f(u)}{C} - \frac{i}{C}; \\ \frac{di}{dt} &= \frac{u}{L}, \end{aligned} \quad (2)$$

где i – ток индуктивности; $f(u)$ – вольт-амперная характеристика (ВАХ) нелинейной проводимости [7].

Точное стационарное решение системы ОДУ (2) и методика его получения для случая, когда дифференциальная проводимость $G(u) = df/du$ описывается выражением (1), представлены в работах [7, 8]. Представленная выше модель автоколебательной цепи позволяет устанавливать требуемый режим работы автогенератора путем изменения параметров G_0 , G_1 и $\sigma = \sqrt{C/L}$, кроме того, для представленной модели можно определить установившиеся значения амплитуды A_{m0} и частоты ω_0 колебаний с точностью до последнего знака разрядной сетки компьютера. В частности, при $L = 1$ Гн и $C = 1$ Ф получены следующие значения амплитуды и частоты [9]:

- ♦ для режима гармонических колебаний $\sigma/G_0 = \sigma/G_1 = 100$
 $A_{m0} = 2,475416990116814$ В, $\omega_0 = 0,99999565970188$ рад/с;
- ♦ для режима релаксационных колебаний $\sigma/G_0 = 0,25$, $\sigma/G_1 = 0,01$
 $A_{m0} = 1,084334602144661$ В, $\omega_0 = 0,345968425746775$ рад/с.

Результаты компьютерного моделирования рассматриваемой автоколебательной цепи в симуляторе Multisim, представленные на рис. 4 и 5 [9], подтверждают низкую эффективность метода Гира для анализа осциллирующих систем, а метода трапеций – для жестких систем.

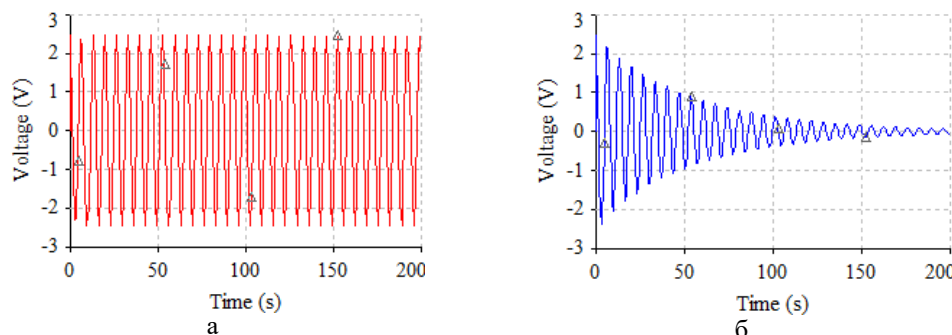


Рис. 4. Результаты численного анализа автоколебательной цепи в режиме гармонических колебаний, полученные методами трапеций (а) и Гира (б) [9]

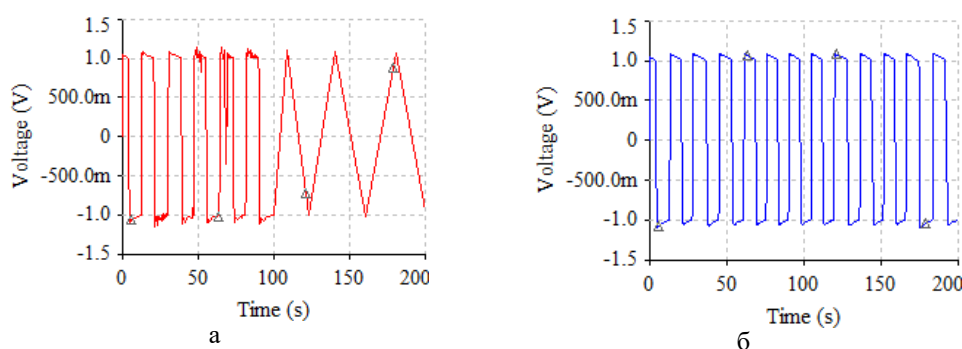


Рис. 5. Результаты численного анализа автоколебательной цепи в режиме релаксационных колебаний, полученные методами трапеций (а) и Гира (б) [9]

2. Разностные схемы численных методов решения ОДУ для SPICE-симуляторов. Рассмотрим численные методы, использующиеся в SPICE-симуляторах, для решения систем ОДУ вида

$$dx/dt = f(x, t), \quad x(0) = x_0,$$

где x – вектор неизвестных переменных, $f(x, t)$ – вектор функция, определяющаяся схемой цепи [11].

Методы Гира 1-го и 2-го порядков ($BDF1$ и $BDF2$) описываются следующими уравнениями (разностными схемами) [12]:

$$x_{n+1} = x_n + hf(x_{n+1}, t_{n+1}); \quad (3)$$

$$x_{n+2} = \frac{4}{3}x_{n+1} - \frac{1}{3}x_n + \frac{2}{3}hf(x_{n+2}, t_{n+2}). \quad (4)$$

где x_{n+2} , x_{n+1} , x_n – приближенные решения системы ОДУ при $t_{n+2} = t_n + 2h$, $t_{n+1} = t_n + h$, $t_n = nh$ соответственно; h – шаг интегрирования; $n = 0, 1, 2, \dots$

Разностная схема метода трапеций (TR) имеет вид [13]:

$$x_{n+1} = x_n + \frac{h}{2}[f(x_n, t_n) + f(x_{n+1}, t_{n+1})]. \quad (5)$$

Следует отметить, что метод Гира 1-го порядка является простейшим неявным методом численного интегрирования ОДУ, также называемым неявным методом Эйлера.

Для преодоления описанных выше недостатков методов Гира (3), (4) и метода трапеций (5) в данной работе для анализа автоколебательных цепей предлагается применять гибридный метод (HM), впервые представленный в работе [14], который показал высокую эффективность численного анализа линейных моделей как осциллирующих систем, так и жестких. Гибридный метод представляет собой комбинацию метода Гира 1-го порядка и метода трапеций:

$$\begin{cases} \mathbf{x}_{n+\alpha} = \mathbf{x}_n + \alpha h \mathbf{f}(\mathbf{x}_{n+\alpha}, t_{n+\alpha}); \\ \mathbf{x}_{n+1} = \mathbf{x}_{n+\alpha} + \frac{(1-\alpha)h}{2} [\mathbf{f}(\mathbf{x}_{n+1}, t_{n+1}) + \mathbf{f}(\mathbf{x}_{n+\alpha}, t_{n+\alpha})], \end{cases} \quad (6)$$

где $\mathbf{x}_{n+\alpha}$ – приближенное решение системы ОДУ при $t_{n+\alpha} = t_n + \alpha h$; $0 \leq \alpha \leq 1$ – весовой коэффициент.

В гибридном методе каждый шаг интегрирования включает в себя два частичных шага, соотношение между которыми определяется весовым коэффициентом α , который представляет собой функцию $\alpha^*(h)$, зависящую от шага интегрирования и удовлетворяющую следующим условиям:

$\alpha^*(h) = 1$ при максимальном шаге интегрирования $h = h_{\max}$, тогда в разностной схеме гибридного метода наибольший вес будет иметь L-устойчивый метод, обладающий наименьшей погрешностью при больших шагах;

$\alpha^*(h) = 0$ при шаге интегрирования $h \rightarrow 0$, тогда в разностной схеме гибридного метода наибольший вес будет иметь Р-устойчивый метод, обладающий наименьшей погрешностью при малых шагах.

Аналитические выражения для функции $\alpha^*(h)$ представлены в работах [15–17]. В данной работе используется функция позволяющая учесть жесткости задачи m для минимизации погрешности численного решения [17]:

$$\alpha = \alpha^*(h) = 1 - [1 - (h/h_{\max})]^m, \quad (7)$$

где $m \approx \lg(\tau_{\max}/\tau_{\min})$; $\tau_{\max}$ и $\tau_{\min}$ – максимальная и минимальная постоянные времени цепи.

График функции (7) показан на рис. 6, из которого следует, что с увеличением жесткости задачи коэффициент α будет быстрее приближаться к единице, при этом в гибридном методе наибольший вес будет иметь L-устойчивый метод.

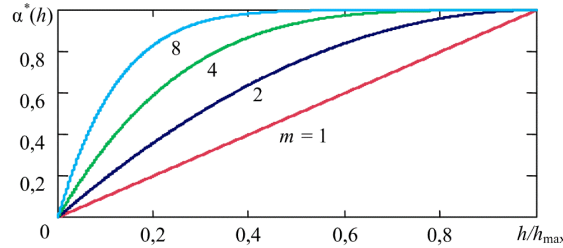


Рис. 6. Зависимости весового коэффициента гибридного метода от нормированного шага интегрирования при различной жесткости задачи

3. Сравнительный анализ точности численных методов. Для количественной оценки точности представленных выше численных методов была использована методика, описанная в работе [4]. В соответствии с данной методикой были определены зависимости относительных погрешностей расчета амплитуды ε_A и частоты ε_ω генерируемых колебаний от шага интегрирования:

$$\varepsilon_A = |A_{mk} - A_{m0}|/A_{m0}; \quad \varepsilon_\omega = |\omega_k - \omega_0|/\omega_0,$$

где A_{mk} и ω_k – оценки амплитуды и частоты колебаний, полученные численными методами.

Зависимости ε_A и ε_ω от шага интегрирования, полученные в различных режимах работы для обобщенной модели автоколебательной цепи представлены на рис. 7 и 8.

Из рис. 7 видно, что в режиме гармонических колебаний гибридный метод обеспечивает примерно такую же точность, как и Р-устойчивый метод трапеций, а погрешности оценки амплитуды и частоты с помощью гибридного метода оказываются значительно меньше, чем у L-устойчивых методов Гира. В частности, для случаев, когда значения отношений шага интегрирования к периоду колебаний составляют $h/T = 10^{-1}$, 10^{-2} и 10^{-3} погрешности гибридного метода будут меньше аналогичных погрешностей метода Гира 2-го порядка примерно в 200, 500 и 30 раз соответственно при оценке амплитуды и примерно в 3, 3,5 и 4 раза соответственно при оценке частоты.

Из рис. 8 видно, что в режиме релаксационных колебаний гибридный метод обеспечивает более высокую точность, чем Р-устойчивый метод трапеций и чем L-устойчивый метод Гира, как при оценке амплитуды, так и при оценке частоты. В частности, для случаев, когда значения отношений $h/T = 10^{-1}$, 10^{-2} и 10^{-3} погрешности гибридного метода будут меньше аналогичных погрешностей метода Гира 2-го порядка примерно в 4, 8 и 40 раз соответственно при оценке амплитуды и примерно в 8, 12 и 6 раз соответственно при оценке частоты, а также меньше аналогичных погрешностей метода трапеций примерно в 10, 20 и 4 раза соответственно при оценке амплитуды и примерно в 10, 3 и 1,5 раза соответственно при оценке частоты.

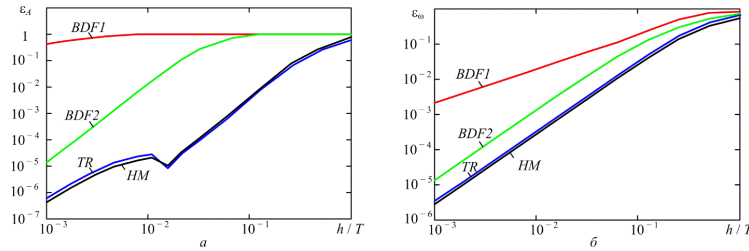


Рис. 7. Относительные погрешности оценки амплитуды (а) и частоты (б) колебаний при численном анализе во временной области автоколебательной цепи в режиме гармонических колебаний

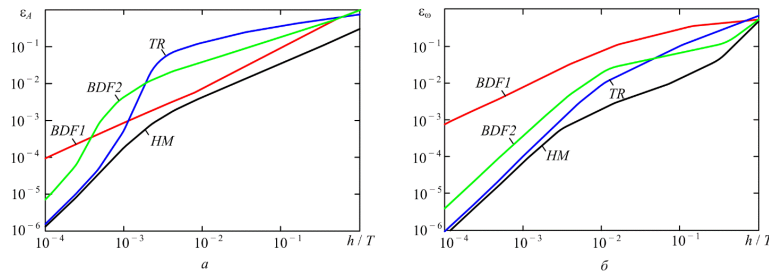


Рис. 8. Относительные погрешности оценки амплитуды (а) и частоты (б) колебаний при численном анализе во временной области автоколебательной цепи в режиме релаксационных колебаний

4. Перспективы развития гибридных методов численного решения ОДУ. Дальнейшее повышение точности и устойчивости гибридных методов возможно за счет применения различных классов неявных методов Рунге-Кутты (IRKM) высокого порядка точности.

Разностная схема метода Рунге-Кутты в общем случае описывается следующими уравнениями [18]:

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{x}_n + h \sum_{i=1}^s b_i \mathbf{f}(\mathbf{X}_i, t_n + c_i h); \\ \mathbf{X}_i &= \mathbf{x}_n + h \sum_{j=1}^s a_{ij} \mathbf{f}(\mathbf{X}_j, t_n + c_j h), \quad 1 \leq i \leq s, \end{aligned} \quad (8)$$

где $\mathbf{X}_i$ – приближенное решения системы ОДУ в момент времени $t_n + c_i h$; s – число стадий метода; $0 \leq c_i \leq 1$, a_{ij} , b_i – вещественные коэффициенты.

Коэффициенты методов Рунге-Кутты представляют в виде таблицы Бутчера:

$$\begin{array}{c|cc} \mathbf{c} & \mathbf{A} & \\ \hline & \mathbf{b}^T & \end{array} = \begin{array}{c|ccc} c_1 & a_{11} & a_{12} & \cdots & a_{1s} \\ c_2 & a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ c_s & a_{s1} & a_{s2} & \cdots & a_{ss} \\ \hline & b_1 & b_2 & \cdots & b_s \end{array}. \quad (9)$$

Наименьшую погрешность численного решения и максимальную L-устойчивость при заданном порядке точности позволяют обеспечить полностью неявные методы Рунге-Кутты (*FIRKM*), разностные схемы которых описаны в работе [19]. Однако данные методы имеют наибольшую вычислительную трудоемкость среди всех неявных методов, поскольку для реализации *FIRKM* порядка p требуется $s = p$ стадий решения системы уравнений (8). Кроме того, *FIRKM* не обладают Р-устойчивостью.

Для понижения вычислительной трудоемкости неявных методов Рунге-Кутты при сохранении порядка точности в работе [20] предлагается использовать подклассы *IRKM* – методы Радо-IIА и Лобатто-IIIА. Методы Радо-IIА являются L-устойчивыми и требуют $s = (p + 1)/2$ вычислительных стадий, а методы Лобатто-IIIА обеспечивают Р-устойчивость и для них достаточно $s = p/2$ вычислительных стадий.

Таблицы Бутчера методов Радо-IIА 1-го, 3-го и 5-го порядков:

			$4 - \sqrt{6}$	$88 - 7\sqrt{6}$	$296 - 169\sqrt{6}$	$-2 + 3\sqrt{6}$
			10	360	1800	225
$\begin{array}{c c} 1 & 1 \\ \hline & 1 \end{array}$	$\frac{1}{3}$	$\begin{array}{c c} 5 & -1 \\ \hline 12 & 12 \end{array}$	$4 + \sqrt{6}$	$296 + 169\sqrt{6}$	$88 + 7\sqrt{6}$	$-2 - 3\sqrt{6}$
	1	$\begin{array}{c c} 3 & 1 \\ \hline 4 & 4 \end{array}$	10	1800	360	225
		$\begin{array}{c c} 3 & 1 \\ \hline 4 & 4 \end{array}$	1	$\frac{16 - \sqrt{6}}{36}$	$\frac{16 + \sqrt{6}}{36}$	$\frac{1}{9}$
				$\frac{16 - \sqrt{6}}{36}$	$\frac{16 + \sqrt{6}}{36}$	$\frac{1}{9}$

Таблицы Бутчера методов Лобатто-IIIА 2-го, 4-го и 6-го порядков:

			0	0	0	0	0
			$5 - \sqrt{5}$	$11 + \sqrt{5}$	$25 - \sqrt{5}$	$25 - 13\sqrt{5}$	$-1 + \sqrt{5}$
			10	120	120	120	120
$\begin{array}{c c} 0 & 0 \\ \hline 1 & 1 \\ \hline 2 & 2 \end{array}$	$\frac{1}{2}$	$\begin{array}{c c} 5 & 1 \\ \hline 24 & 3 \end{array}$	$5 + \sqrt{5}$	$11 - \sqrt{5}$	$25 + 13\sqrt{5}$	$25 + \sqrt{5}$	$-1 - \sqrt{5}$
	1	$\begin{array}{c c} 1 & 2 \\ \hline 6 & 3 \end{array}$	10	120	120	120	120
		$\begin{array}{c c} 1 & 2 \\ \hline 6 & 3 \end{array}$	1	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{12}$
				$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{12}$

Следует отметить, что метод Гира 1-го порядка представляет собой метод Радо IIА 1-го порядка, а метод трапеций – метод Лобатто IIIА 2-го порядка. Таким образом, разностную схему гибридного метода высокого порядка точности можно получить путем замены в разностной схеме (6) выражения для первого частичного шага $\mathbf{x}_{n+\alpha}$ обобщенной разностной схемой метода Радо IIА, а выражения для второго частичного шага $\mathbf{x}_{n+1}$ – обобщенной разностной схемой метода Лобатто IIIА:

$$\begin{cases} \mathbf{X}\mathbf{R}_i = \mathbf{x}_n + ah \sum_{j=1}^{sr} ar_{ij} \mathbf{f}(\mathbf{X}\mathbf{R}_j, t_n + cr_j ah), \\ \mathbf{x}_{n+\alpha} = \mathbf{x}_n + ah \sum_{i=1}^{sr} br_i \mathbf{f}(\mathbf{X}\mathbf{R}_i, t_n + cr_i ah); \\ \mathbf{X}\mathbf{L}_i = \mathbf{x}_{n+\alpha} + (1 - \alpha)h \sum_{j=1}^{sl} al_{ij} \mathbf{f}(\mathbf{X}\mathbf{L}_j, t_n + cl_j(1 - \alpha)h), \\ \mathbf{x}_{n+1} = \mathbf{x}_{n+\alpha} + (1 - \alpha)h \sum_{i=1}^{sl} bl_i \mathbf{f}(\mathbf{X}\mathbf{L}_i, t_n + cl_i(1 - \alpha)h), \end{cases} \quad (10)$$

где ar_{ij} , br_i , cr_i и al_{ij} , bl_i , cl_i – коэффициенты методов Радо IIА и Лобатто IIIА, представленные в таблицах Бутчера; sr и sl – число стадий в методах Радо IIА и Лобатто IIIА.

Выводы. Предлагаемый гибридный метод позволяет обеспечить наименьшую погрешность моделирования автоколебательных цепей, как в режиме гармонических колебаний, так и в режиме релаксационных колебаний. Для получения максимальной эффективности гибридного метода важно корректно выбирать весовой коэффициент, опреде-

ляющий соотношение между частичными шагами численных методов, входящих в разностную схему гибридных методов. При повышении жесткости задачи, определяющейся отношением максимальной и минимальной постоянных времени цепи, необходимо увеличивать степень функции, описывающей весовой коэффициент примерно до значения примерно равного степени жесткости задачи.

Дальнейшее повышение точности и устойчивости гибридного метода возможно за счет применения неявных методов Рунге-Кутты высокого порядка. Однако применение неявных методов Рунге-Кутты высокого порядка может приводить к существенному повышению вычислительной сложности гибридного метода, получаемого на их основе, поэтому необходимо проводить дополнительные исследования эффективности при разработке новых гибридных методов с учетом вычислительной трудоемкости.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Энгельхард М., Ревенко А.* SPICE Differentiation. Различия в реализациях симуляторов SPICE // Компоненты и технологии. – 2015. – № 7 (168). – С. 82-88.
2. *Прикота А.* Программа SimOne. Современные средства схемотехнического моделирования // Электроника НТБ. – 2011. – № 2. – С. 142-146.
3. *Petzold L.R., Jay L.O., Yen J.* Numerical solution of highly oscillatory ordinary differential equations // Acta Numerica. – 1997. – P. 437-483.
4. *Guglielmi N., Hairer E.* Applying Stiff Integrators for Ordinary Differential Equations and Delay Differential Equations to Problems with Distributed Delays // SIAM Journal on Scientific Computing. – 2025. – Vol. 47, No. 1. – DOI: 10.1137/24M1632413.
5. *Жук Д.М., Кожевников Д.Ю., Маничев В.Б.* Проблемы разработки математического ядра для программ моделирования динамики технических систем // Проблемы разработки перспективных микро- и наноэлектронных систем (МЭС). – 2020. – № 4. – С. 31-38. – DOI: 10.31114/2078-7707-2020-4-31-38.
6. *Bernardini A., Maffezzoni P., Sarti A.* Linear Multistep Discretization Methods With Variable Step-Size in Nonlinear Wave Digital Structures for Virtual Analog Modeling // IEEE/ACM Transactions on Audio, Speech, and Language Processing. – 2019. – Vol. 27, No. 11. – P. 1763-1776. – DOI: 10.1109/TASLP.2019.2931759.
7. *Пилипенко А.М., Бирюков В.Н.* Исследование эффективности современных численных методов при анализе автоколебательных цепей // Журнал радиоэлектроники. – 2013. – № 8. – URL: <http://jre.cplire.ru/jre/aug13/9/text.html> (дата обращения 24.08.2025).
8. *Пилипенко А.М., Бирюков В.Н.* Оценка точности численного анализа релаксационного генератора // Журнал радиоэлектроники. – 2013. – № 11. – С. 7. – URL: <http://jre.cplire.ru/jre/nov13/6/text.html> (дата обращения 24.08.2025).
9. *Пилипенко А.М., Агабекян А.В.* Модель автоколебательной цепи для тестирования методов численного анализа переходных процессов в SPICE-симуляторах // Известия ЮФУ. Технические науки. – 2022. – № 3(227). – С. 243-254. – DOI: 10.18522/2311-3103-2022-3-243-254.
10. *Бирюков В.Н.* Руководство к лабораторным работам по курсу «Основы автоматизированного анализа цепей». – Вып. 3. – Таганрог: Изд-во ТРТУ, 2001. – 27 с.
11. *Пантелеев А.В., Якимов А.С., Босов А.В.* Обыкновенные дифференциальные уравнения: основы теории и алгоритмы решений: учеб. пособие. – М.: Вузовская книга, 2012. – 188 с.
12. *Gear C.W.* Algorithm 407: DIFSUB for Solution of Ordinary Differential Equation // Communications of the ACM. – 1971. – Vol. 14, No. 4. – P. 185-190.
13. *Чуа Л.О. Пен-Мин Л.* Машинный анализ электронных схем: алгоритмы и вычислительные методы: пер. с англ. Е.С. Виленкина и др. / под ред. В.Н. Ильина. – М.: Энергия, 1980. – 640 с.
14. *Pilipenko A.M., Biryukov V.N.* Development and testing of high accuracy hybrid methods for time-domain simulation of electronic circuits and systems // Proceedings of 2017 IEEE East-West Design and Test Symposium, EWDTs. – 2017. – P. 8110124. – DOI 10.1109/EWDTs.2017.8110124.
15. *Бирюков В.Н., Пилипенко А.М.* Численный анализ жестких узкополосных систем // Радиосистемы. – 2002. – № 2 (62). – С. 36-38.
16. *Пилипенко А.М. Бирюков В.Н.* Гибридные методы решения обыкновенных дифференциальных уравнений жестких и/или колебательных цепей // Радиотехника. – 2011. – № 1. – С. 11-15.
17. *Пилипенко А.М.* Гибридные методы высокого порядка точности для численного анализа во временной области жестких и колебательных цепей // Моделирование, оптимизация и информационные технологии. – 2017. – № 3 (18). – <https://moitvvt.ru/ru/journal/pdf?id=367> (дата обращения: 24.08.2025).

18. Хайпер Э. Ваннер Г. Решение обыкновенных дифференциальных уравнений: Жесткие и дифференциально-алгебраические задачи: пер. со второго англ. изд. Е.Л. Старостина и др. / под ред. С.С. Филиппова. – М.: Мир, 1999. – 685 с.
19. Ракитский Ю.В., Устинов С.М., Черноуцкий И.Г. Численные методы решения жестких систем. – М.: Наука, 1979. – 208 с.
20. Maffezzoni P., Codecasa L., D'Amore D. Time-domain simulation of nonlinear circuits through implicit Runge-Kutta methods // IEEE Transactions. Circuits and Systems. – 2007. – Vol. 54, No. 2. – P. 391-400.

REFERENCES

1. Engel'khard M., Revenko A. SPICE Differentiation. Razlichiya v realizatsiyakh simulyatorov SPICE [SPICE Differentiation. Differences in SPICE Simulator Implementations], *Komponenty i tekhnologii* [Components and Technologies], 2015, No. 7 (168), pp. 82-88.
2. Prikota A. Programma SimOne. Sovremennyye sredstva skhemotekhnicheskogo modelirovaniya [SimOne Program. Modern Means of Circuit Simulation], *Elektronika NTB* [Electronics Science Technology Business], 2011, No. 2, pp. 142-146.
3. Petzold L.R., Jay L.O., Yen J. Numerical solution of highly oscillatory ordinary differential equations, *Acta Numerica*, 1997, pp. 437-483.
4. Guglielmi N., Hairer E. Applying Stiff Integrators for Ordinary Differential Equations and Delay Differential Equations to Problems with Distributed Delays, *SIAM Journal on Scientific Computing*, 2025, Vol. 47, No. 1. DOI: 10.1137/24M1632413.
5. Zhuk D.M., Kozhevnikov D.Yu., Manichev V.B. Problemy razrabotki matematicheskogo yadra dlya programm modelirovaniya dinamiki tekhnicheskikh sistem [Problems of developing a mathematical kernel for programs for modeling the dynamics of technical systems], *Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh sistem (MES)* [Problems of developing promising micro- and nanoelectronic systems (MES)], 2020, No. 4, pp. 31-38. DOI: 10.31114/2078-7707-2020-4-31-38.
6. Bernardini A., Maffezzoni P., Sarti A. Linear Multistep Discretization Methods With Variable Step-Size in Nonlinear Wave Digital Structures for Virtual Analog Modeling, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2019, Vol. 27, No. 11, pp. 1763-1776. DOI: 10.1109/TASLP.2019.2931759.
7. Pilipenko A.M., Biryukov V.N. Issledovanie effektivnosti sovremennykh chislennykh metodov pri analize avtokolebatel'nykh tsepey [Study of the efficiency of modern numerical methods in the analysis of self-oscillating circuits], *Zhurnal radioelektroniki* [Journal of Radio Electronics], 2013, No. 8. Available at: <http://jre.cplire.ru/jre/aug13/9/text.html> (accessed 24 August 2025).
8. Pilipenko A.M., Biryukov V.N. Otsenka tochnosti chislennogo analiza relaksatsionnogo generatora [Estimation of the accuracy of numerical analysis of a relaxation oscillator], *Zhurnal radioelektroniki* [Zhurnal Radioelektroniki], 2013, No. 11, pp. 7. Available at: <http://jre.cplire.ru/jre/nov13/6/text.html> (accessed 24 August 2025).
9. Pilipenko A.M., Agabekyan A.V. Model' avtokolebatel'noy tsepi dlya testirovaniya metodov chislennogo analiza perekhodnykh protsessov v SPICE-simulyatorakh [Model of a self-oscillating circuit for testing methods of numerical analysis of transient processes in SPICE simulators], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2022, No. 3 (227), pp. 243-254. DOI: 10.18522/2311-3103-2022-3-243-254.
10. Biryukov V.N. Rukovodstvo k laboratornym rabotam po kursu «Osnovy avtomatizirovannogo analiza tsepey» [Laboratory Manual for the Course "Fundamentals of automated circuit analysis"], Issue 3. Taganrog: Izd-vo TRTU, 2001, 27 p.
11. Panteleev A.V., Yakimova A.S., Bosov A.V. Obyknoennyye differentsial'nye uravneniya: osnovy teorii i algoritmy resheniy: ucheb. posobie [Ordinary differential equations: fundamentals of theory and solution algorithms: textbook]. Moscow: Vuzovskaya kniga, 2012, 188 p.
12. Gear C.W. Algorithm 407: DIFSUB for Solution of Ordinary Differential Equation, *Communications of the ACM*, 1971, Vol. 14, No. 4, p. 185-190.
13. Chua L.O., Pen-Min L. Mashinnyy analiz elektronnykh skhem: algoritmy i vychislitel'nye metody [Machine analysis of electronic circuits: algorithms and computational methods]: transl. from Engl. by E.S. Vilenkina i dr., ed. by V.N. Il'ina. Moscow: Energiya, 1980, 640 p.
14. Pilipenko A.M., Biryukov V.N. Development and testing of high accuracy hybrid methods for time-domain simulation of electronic circuits and systems, *Proceedings of 2017 IEEE East-West Design and Test Symposium, EWDTS*, 2017, pp. 8110124. DOI 10.1109/EWDTS.2017.8110124.
15. Biryukov V.N., Pilipenko A.M. Chislennyy analiz zhestkikh uzkopolosnykh sistem [Numerical analysis of rigid narrowband systems], *Radiosistemy* [Radio Systems], 2002, No. 2 (62), pp. 36-38.

16. Pilipenko A.M., Biryukov V.N. Gibridnye metody resheniya obyknovennykh differentsial'nykh uravneniy zhestkikh i/ili kolebatel'nykh tsepey [Hybrid methods for solving ordinary differential equations of rigid and/or oscillatory circuits], *Radiotekhnika* [Radio Engineering], 2011, No. 1, pp. 11-15.
17. Pilipenko A.M. Gibridnye metody vysokogo poriadka tochnosti dlya chislennogo analiza vo vremennoy oblasti zhestkikh i kolebatel'nykh tsepey [Hybrid methods of high order of accuracy for numerical analysis in the time domain of rigid and oscillatory circuits], *Modelirovanie, optimizatsiya i informatsionnye tekhnologii* [Modeling, Optimization and Information Technology], 2017, No. 3 (18). Available at: <https://moitvvt.ru/ru/journal/pdf?id=367> (accessed 24 August 2025).
18. Khayrer E., Vanner G. Reshenie obyknovennykh differentsial'nykh uravneniy: Zhestkie i differentsial'no-algebraicheskie zadachi [Solution of ordinary differential equations: Stiff and differential-algebraic problems], transl. from the second Engl. ed. by E.L. Starostina i dr., ed. by S.S. Filippova. Moscow: Mir, 1999, 685 p.
19. Rakitskiy Yu.V., Ustinov S.M., Chernorutskiy I.G. Chislennye metody resheniya zhestkikh system [Numerical methods for solving stiff systems]. Moscow: Nauka, 1979, 208 p.
20. Maffezzoni P., Codecasa L., D'Amore D. Time-domain simulation of nonlinear circuits through implicit Runge-Kutta methods, *IEEE Transactions. Circuits and Systems*, 2007, Vol. 54, No. 2, pp. 391-400.

Пилипенко Александр Михайлович – Южный федеральный университет; e-mail: ampilipenko@sfedu.ru; г. Таганрог, Россия; тел.: +78634371632; кафедра теоретических основ радиотехники; зав. кафедрой; к.т.н.; доцент.

Pilipenko Alexandr Mikhailovich – Southern Federal University; e-mail: ampilipenko@sfedu.ru; Taganrog, Russia; phone: +78634371632; the Department of Fundamentals of Radio Engineering; head of the department; cand. of eng. sc.; associate professor.

УДК 004.89 + 004.042

DOI 10.18522/2311-3103-2026-1-122-147

В.В. Гапочка, Е.Е. Полупанова

ОБЗОР ТЕХНОЛОГИЙ СЕМАНТИЧЕСКОГО АНАЛИЗА И ИНТЕГРАЦИИ ГЕТЕРОГЕННЫХ ИНФОРМАЦИОННЫХ ПОТОКОВ В СИСТЕМАХ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

Современные системы поддержки принятия решений (СППР) работают с потоками данных из IoT-сенсоров, баз данных, текстовых сообщений и социальных медиа, представленных в различных форматах и обладающих неодинаковыми смысловыми моделями и качеством. Отсутствие семантически согласованной интеграции приводит к неоднозначной интерпретации сущностей, дублированию и потере контекста, что снижает качество и оперативность решений. Цель работы – систематизировать методы семантического анализа и интеграции гетерогенных информационных потоков для СППР, выявить их преимущества, ограничения и области применения. Исследование выполнено в форме аналитического обзора публикаций 2018–2025 гг. по семантической интероперабельности, онтологиям и знаниевым графам, многоисточниковому слиянию данных, федеративному доступу и потоковой обработке. Показано, что обеспечение семантической совместимости достигается за счёт онтологий и знаниевых графов, задающих общие сущности и идентификаторы и обеспечивающих гибкость схемы интеграции. Для режимов реального времени наиболее результативны гибридные решения, сочетающие семантический слой с алгоритмами слияния и учётом доверия к источникам; в прикладных кейсах отмечаются улучшения точности порядка 15–20% и сокращение времени отклика до 70–80% при использовании многоисточниковых моделей. Для неструктурированных потоков ключевую роль играют NLP и машинное обучение, обеспечивающие извлечение сущностей и отношений и семантическое обогащение данных. Полученные результаты могут быть использованы при проектировании СППР в доменах умного города, промышленности и здравоохранения. Кроме того, работа подчеркивает важность стандартов, таких как SHACL для валидации и SPARQL для запросов, которые усиливают практическую применимость семантических подходов. Перспективы включают автоматизацию ontology alignment для снижения трудозатрат и интеграцию с ИИ для динамической адаптации к новым источникам данных.

Семантическая интеграция данных; гетерогенные информационные потоки; система поддержки принятия решений; онтология; знаниевый граф; семантическая интероперабельность; слияние данных; федерация данных; обработка потоков; интернет вещей.

V.V. Gapochka, E.E. Polupanova

**SEMANTIC ANALYSIS AND INTEGRATION OF HETEROGENEOUS
INFORMATION STREAMS IN DECISION SUPPORT SYSTEMS:
A TECHNOLOGY REVIEW**

Modern decision support systems (DSS) increasingly rely on heterogeneous data streams from IoT sensors, databases, text messages, and social media, represented in different formats and characterized by diverse semantic models and quality levels. The lack of semantically aligned integration results in inconsistent entity interpretation, duplication, and loss of context, which reduces the quality and timeliness of decisions. The aim of this paper is to systematize methods for semantic analysis and integration of heterogeneous information streams in DSS and to identify their benefits, limitations, and application domains. The study is conducted as an analytical review of publications from 2018–2025 focusing on semantic interoperability, ontologies and knowledge graphs, multi-source data fusion, data federation, and real-time stream processing. The review shows that semantic compatibility is primarily achieved through ontologies and knowledge graphs that define shared entities and identifiers and provide a flexible integration schema. For real-time decision-making, hybrid solutions combining a semantic layer with data fusion algorithms and source trust assessment are the most effective; published case studies report accuracy gains of about 15–20% and response-time reductions of up to 70–80% in multi-source settings. For unstructured streams, NLP and machine learning play a key role by extracting entities and relations and enabling semantic enrichment. The results can be used to design DSS for smart city, industrial, and healthcare domains. Furthermore, the paper highlights the role of standards like SHACL for validation and SPARQL for querying, enhancing the practical applicability of semantic approaches. Future directions include automating ontology alignment to reduce labor costs and integrating with AI for dynamic adaptation to new data sources.

Semantic data integration; heterogeneous information streams; decision support system; ontology; knowledge graph; semantic interoperability; data fusion; data federation; stream processing; Internet of Things.

Введение. В эпоху больших данных и интернета вещей организации и города сталкиваются с огромными потоками разнородной информации, которые необходимо учитывать при принятии решений. Гетерогенные информационные потоки включают структурированные данные (таблицы БД), полуструктурированные (XML/JSON документы, лог-файлы) и неструктурированные данные (текстовые отчеты, изображения, датчики, посты соцсетей). Каждый тип данных обладает своими форматами, временными характеристиками и семантическими представлениями – словарями и контекстом, в рамках которых данные имеют смысл. Традиционные СППР, как правило, разрабатываются под один источник или тип данных и не обладают встроенными средствами унификации информации из разных источников. В результате актуальной становится проблема семантической интероперабельности: данные из разных систем и датчиков могут иметь разный смысл или структуру, что приводит к ошибкам интерпретации, дублированию и пробелам в знании.

Актуальность проблемы. В научной и прикладной сфере назрел запрос на решения, обеспечивающие целостное представление ситуации на основе мультимедийных данных. Например, в умном городе для управления дорожным движением нужны данные от транспортных датчиков, погодных служб, социальных сетей (сообщения о ДТП) и городских реестров. Эти сведения разнородны по природе – числовые показания, текстовые сообщения, ГИС-данные – и поступают с различной периодичностью (потоково и дискретно). Без их интеграции городские службы не смогут получить полную картину и своевременно реагировать на события. Аналогичные вызовы существуют в медицине (совмещение электронных медкарточек, результатов анализов и научных публикаций), в промышленности (данные с разных участков производства), в финансах (агрегирование новостей, котировок, экономических индикаторов) и т.д. Везде правильность решений зависит от способности системы обработать весь спектр доступной информации. Однако сегодня многие информационные системы остаются изолированными и разработаны под узкие задачи, что затрудняет их совместное использование. Даже при наличии технической интеграции на уровне форматов (синтаксическая совместимость) остается проблема семантической несовместимости – разных терминологий, онтологий и контекстов данных в разных источниках.

Анализ текущего состояния. В последние пять лет наблюдается интенсивное развитие методов и инструментов для интеграции гетерогенных данных [1–4]. Согласно обзору 2024 г. [3], большинство исследований сфокусировано на семантических аспектах объединения данных, особенно на проблеме интеграции неструктурированных данных и разрешении различий в терминологии. Это объясняется тем, что именно семантические несоответствия (разные названия одних сущностей, разные классификационные схемы, единицы измерения и т.д.) являются главным препятствием к полноценной интеграции [3]. Появились специализированные решения: разработаны онтологии для различных доменов (например, онтологии сенсорных сетей, медицинские онтологии), создаются знаниевые графы, охватывающие множество источников, и разрабатываются алгоритмы выравнивания схем данных (schema matching) на основе правил и машинного обучения. С другой стороны, эволюция технологий больших данных привела к тому, что интеграция осуществляется на двух уровнях: физическом и виртуальном [3]. Физический подход (например, хранилища данных, data warehouse) обеспечивает более эффективные запросы, но дорог в реализации и обновлении; виртуальный подход (посредники, вытягивающие данные по запросу из источников) набирает популярность благодаря гибкости и снижению дублирования данных. Многие современные работы применяют сочетание этих подходов – создание единого семантического слоя, который виртуально объединяет данные из разных хранилищ.

Однако остаются нерешенные вопросы. Во-первых, обеспечение масштабируемости: с ростом числа источников и объема данных возрастает нагрузка на системы интеграции, требуя распределенных архитектур и оптимизированных алгоритмов. Во-вторых, автоматизация семантического согласования: ручная разработка онтологий и правил сопоставления трудоемка, поэтому актуально применение методов ИИ, способных автоматически находить соответствия между понятиями разных датасетов. В-третьих, вопрос качества и доверия к данным: например, данные из социальных сетей ценны для быстрого обнаружения событий, но содержат шум и могут быть недостоверными. Нужно внедрение механизмов верификации, очистки и оценки доверия к источникам (например, учет рейтинга надежности источника в алгоритмах слияния). Наконец, проблема стандартизации: отсутствие единых стандартов описания и передачи данных затрудняет интероперабельность, хотя в отдельных отраслях наблюдается прогресс (например, стандарты HL7 FHIR в здравоохранении, ISO 15926 в промышленности) [4–6].

Цель исследования. С учетом изложенного, целью данной работы является обобщение и систематизация существующих методов семантического анализа и интеграции разнородных информационных потоков для СППР, а также оценка их эффективности. Иными словами, работа отвечает на вопрос: *какие технологии позволяют объединить гетерогенные данные разных типов в единую информационную основу для принятия решений, и какой эффект дает их применение?*

Объект и предмет исследования. Объектом исследования являются процессы и системы интеграции данных в рамках СППР, оперирующие множественными источниками информации. Предмет исследования – методы семантического анализа и интеграции, позволяющие преодолеть структурные и семантические различия данных (включая онтологии, алгоритмы слияния, архитектурные решения). Гипотеза исследования заключается в том, что использование семантических моделей знаний (онтологий, знаниевых графов) в сочетании с автоматизированными методами обработки данных (машинное обучение, потоковая аналитика) значительно повышает полноту и точность информации, доступной СППР, что в итоге повышает качество решений.

В дальнейшем изложении статьи сначала рассматриваются типы гетерогенных данных и связанные с ними проблемы интеграции, затем описываются основные подходы и методы решения поставленной задачи. Далее приводятся примеры реализации интеграции в различных доменах и обсуждаются достигнутые результаты, а в заключении сформулированы основные выводы, новизна и направления дальнейших исследований.

Типы гетерогенных информационных потоков и проблемы их интеграции.

Для проектирования интеграции в СППР целесообразно диагностировать гетерогенность по нескольким осям:

- 1) структура представления данных;
- 2) смысловая (семантическая) модель;
- 3) качество и динамика (задержка, актуальность, шум);
- 4) способы доступа и протоколы.

На практике эти аспекты накладываются: один и тот же объект может одновременно присутствовать в табличных данных, потоковых JSON-событиях и текстовых сообщениях, отличаясь терминологией и уровнем доверия к источнику (табл. 1).

Таблица 1

Матрица типов потоков

	Batch	Near-real-time	Streaming
Structured	DWH/OLAP выгрузки, таблицы БД → ETL/ELT, CDC	Оперативные таблицы, частые инкременты → CDC, upsert	Телеметрия как time-series, агрегаты → window/CEP, merge
Semi-structured	XML/JSON файлы, логи → парсинг, канонизация	API-ответы JSON, schema drift → schema registry, нормализация	JSON events (Kafka), IoT payload → stream ETL, JSON-LD [7]/RDF
Unstructured	Документы/отчёты/PDF → NLP (NER/RE), извлечение фактов	Новости/соцмедиа пачками → NLP	Соцмедиа/чат/аудио/видео → online NLP/CV, KG enrichment

Примечание: CDC – Change Data Capture; CEP – Complex Event Processing; NER/RE – извлечение сущностей/отношений; KG – Knowledge Graph (граф знаний); CV – computer vision.

Далее гетерогенность рассмотрим по ключевым аспектам.

Структурная гетерогенность. Данные различаются по структуре и модели представления. Одни источники могут хранить информацию в виде реляционных таблиц со строгой схемой, другие – в виде иерархических документов (XML/JSON), третьи – как неструктурированный текст или мультимедиа. Например, система мониторинга города содержит табличные данные о транспортных потоках, геопривязанные временные ряды с датчиков (формат JSON) и текстовые сообщения о происшествиях. При объединении таких данных возникает проблема согласования разных схем и форматов. Требуется привести их к единой представленной форме, не потеряв смысл. Часто используют преобразование всех данных к формату RDF (Resource Description Framework) или другим промежуточным моделям, позволяющим описать разную структуру единым способом [8].

Семантическая гетерогенность. Разные источники описывают схожие сущности различными терминами и моделями. Классический пример – различия в онтологиях: в одной базе поле названо “*CitizenID*”, в другой “*PersonCode*”, и, хотя по смыслу это одно и то же, без явного соответствия система воспримет их как разные понятия (рис. 1). Аналогично, единицы измерения (например, градусы Цельсия и Фаренгейта) или классификаторы (разные рубрики товаров в двух магазинах) – все это создает барьеры для автоматического объединения. Семантическая несовместимость ведет к тому, что агрегированные данные могут быть неправильно интерпретированы или не сопоставлены друг с другом. Решение требует введения слоя знаний – словарей, тезаурусов или онтологий, которые задают единую систему понятий для всех источников [9]. Составление такого словаря – сложная методическая задача, особенно если данные охватывают несколько предметных областей сразу.

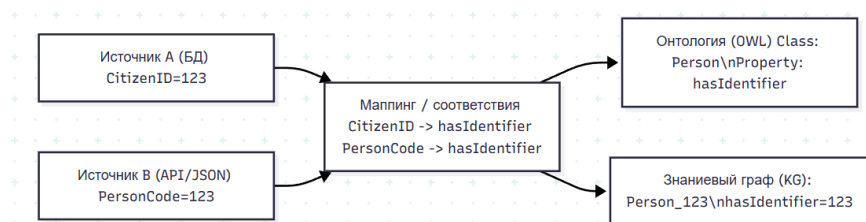


Рис. 1. Пример мэппинга данных из разных источников

Различие в качестве и динамике данных. Информационные потоки различаются по достоверности, точности, частоте обновления. Например, данные от официальных датчиков обычно надежны, но обновляются с интервалом в минуты, тогда как сообщения граждан в соцсетях появляются мгновенно, но могут содержать ошибки или дезинформацию [10]. Интергируя такие источники, СППР должна учитывать вес или доверие к каждому. Возникает задача валидации и фильтрации: выявлять аномалии, пропуски, противоречия между источниками. Необходимо обеспечить, чтобы слияние данных повышало информационную ценность, а не вносило дополнительный шум. Здесь помогают методы фильтрации на основе правил качества данных и статистические модели, оценивающие достоверность (к примеру, кросс-проверка сообщений соцсетей с датчиками или другими независимыми источниками). Современные решения включают также *семантическую валидацию* – проверку соответствия входящего потока заданным онтологиям и ограничениям (например, в умном городе – сравнение показаний датчиков с допустимыми диапазонами и контекстом) [11] (рис. 2).

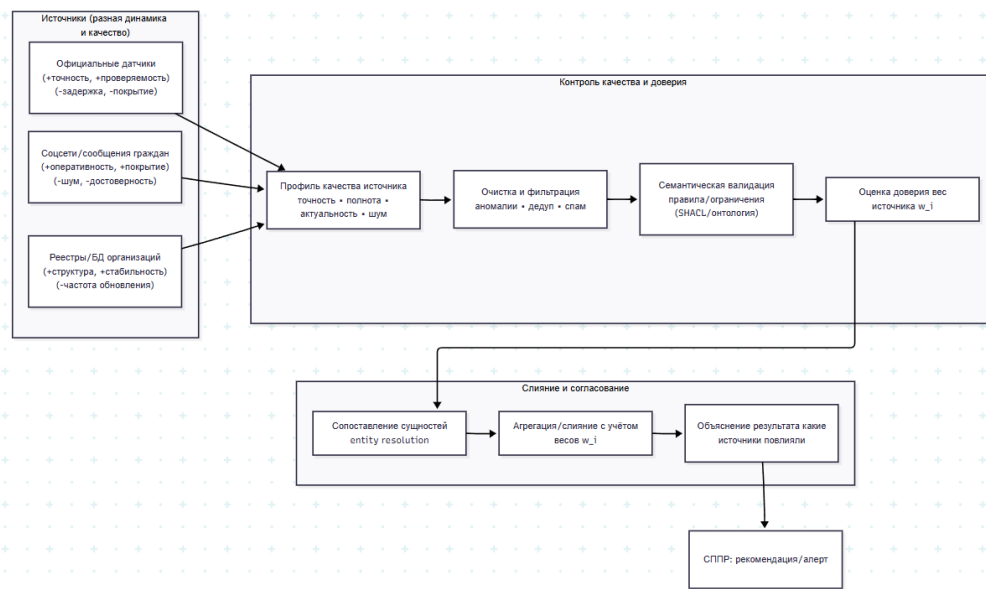


Рис. 2. Контроль качества и учёт доверия к источникам при интеграции потоков данных

Существует большое разнообразие протоколов и форматов доступа. Одни данные доступны через API в реальном времени, другие – как статические файлы, третьи – через базы данных. Интеграционная система должна уметь *подключаться* к разным источникам: читать стримы событий (например, через MQTT или Kafka), выполнять SQL-запросы к БД, парсить веб-страницы и т.д. Это больше технический аспект, но без его решения невозможно собрать все данные воедино. В рамках современных платформ все

чаще применяется микросервисная архитектура, где специализированные коннекторы собирают данные с каждого типа источника и передают их в единый центр хранения или обработки [12] (рис. 3).

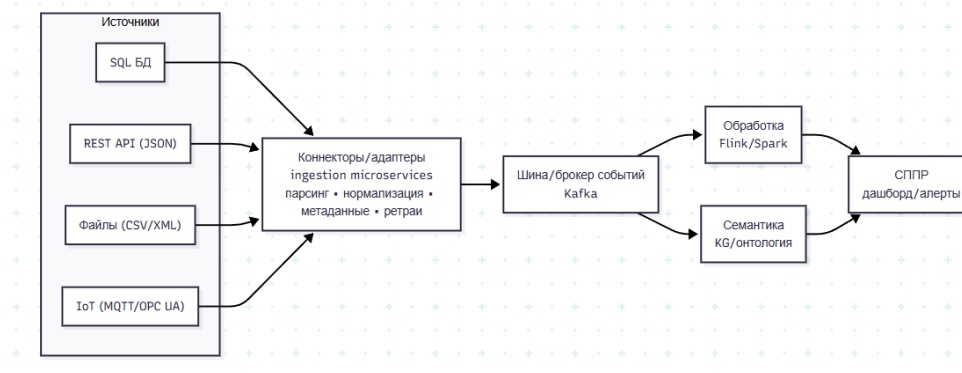


Рис. 3. Слой коннекторов доступа. Унификация протоколов и форматов

Понимание этих видов гетерогенности позволяет сформировать требования к методам интеграции. В частности, необходимы:

- 1) средства трансформации данных между форматами;
- 2) семантические шлюзы – словари или сервисы сопоставления понятий;
- 3) механизмы оценки и обеспечения качества данных;
- 4) модульная архитектура, поддерживающая подключение новых источников.

Методы и подходы к семантической интеграции данных. После описания видов гетерогенности и связанных с ними проблем (структурной, семантической, качественной и гетерогенности доступа) возникает практический вопрос: какими методами и технологиями эти проблемы решаются в современных системах поддержки принятия решений. В отличие от “классической” интеграции данных (ETL/хранилища), семантическая интеграция требует не только приведения форматов, но и согласования смысла, разрешения неоднозначностей сущностей, контроля доверия к источникам и поддержки запросов поверх множества разнородных потоков. Поэтому в современной литературе и практике наблюдается переход к многоуровневым решениям, где семантический слой (онтологии/графы знаний) работает совместно с алгоритмами слияния, средствами потоковой обработки и архитектурными механизмами виртуализации/федерации доступа [3, 12].

Классификация подходов и критерии выбора. Современные подходы к семантической интеграции гетерогенных информационных потоков в СППР целесообразно сгруппировать по доминирующему принципу решения задачи на три класса:

- 1) знаниевые (knowledge-driven);
- 2) алгоритмические (data-/compute-driven);
- 3) архитектурно-гибридные.

Знаниевые подходы ориентированы на согласование смысла данных через онтологии, словари и знаниевые графы, обеспечивая единый “язык” предметной области и сквозную идентификацию сущностей. Алгоритмические подходы концентрируются на вычислительной стороне интеграции: сопоставлении сущностей, слиянии данных, обработке неопределенности и учёте доверия к источникам, включая методы машинного обучения для entity matching и извлечения фактов из неструктурированных потоков. Архитектурно-гибридные решения обеспечивают внедрение этих методов в условиях больших объемов и скоростей: потоковая обработка, микросервисные коннекторы, семантический слой поверх lake/warehouse, а также виртуальная интеграция и федеративные запросы без полного копирования данных.

Выбор конкретного набора методов определяется эксплуатационными требованиями и характеристиками данных. При строгих требованиях к задержке (latency) и реакции в реальном времени приоритет получают потоковые и гибридные решения, где семантический слой

дополняет stream processing и механизмы оценки качества. При высокой доле неструктурированных данных (тексты, изображения, сообщения) критичны NLP/CV и извлечение сущностей/отношений с последующим связыванием с онтологиями и графом знаний. Если для лица, принимающего решение, важна объяснимость (почему система выдала рекомендацию), предпочтительны модели с явным знанием (онтологии, KG) и трассируемыми правилами валидации/качества. При большом числе источников и частых изменениях схем возрастает роль гибких архитектур и механизмов виртуализации/федерации, снижающих стоимость поддержки интеграции. Наконец, ограничения на хранение и копирование данных (организационные, юридические, технические) напрямую влияют на выбор между материализацией интегрированных данных и их виртуальным объединением.

Сводное сравнение подходов приведено в табл. 2.

Таблица 2

Классификация подходов и критерии выбора

Подход	Сильные стороны	Ограничения	Когда применять	Примеры технологий/практик
Знаниевый (онтологии, KG)	Единая смысловая модель; устранение терминологических конфликтов; сквозная идентификация сущностей; поддержка семантических запросов и объяснимости	Трудоёмкость разработки/поддержки онтологий; требования к качеству маппингов; масштабируемость больших графов	Нужно “единое понимание” данных, сильная объяснимость, много разнородных источников и терминологий	RDF/OWL [13]; Knowledge Graph; SPARQL; триплетные хранилища; ontology alignment
Алгоритмический (matching/fusion/ML)	Хорошо работает с шумом и неопределённостью; автоматизация сопоставления записей; повышение точности при мульти-источниковости; применим к потокам	Может быть менее объясним; требует данных для обучения/калибровки; чувствителен к дрейфу данных	Есть конфликтующие источники, нужна оценка доверия, много событий/сенсоров, требуется online-обработка	Data fusion (Bayes/Kalman/DS); entity resolution; ML-matching; NLP-извлечение фактов
Архитектурно-гибридный (stream + семантика + федерация)	Низкая задержка; масштабируемость; гибкое подключение новых источников; возможность не копировать данные (виртуализация/федерация)	Сложность внедрения и эксплуатации; зависимость от зрелости инфраструктуры; возможный рост латентности при федерации	Реальное время, много коннекторов, быстро меняющиеся источники, ограничения на копирование данных	Kafka/Flink/Spark; микросервисы ingestion; Trino/федерация; семантический слой поверх lake/warehouse
Виртуальная интеграция (семантическая федерация)	Актуальные данные “из источника”; минимизация дублирования; снижение затрат на хранение	Латентность запросов; сложность оптимизации; зависимость от доступности источников	Нужны “самые свежие” данные и/или нельзя реплицировать источники	Ontop (R2RML); Trino; query federation; wrappers/mediators
Материализация (ETL/ELT + семантический слой)	Высокая скорость аналитики; устойчивость и воспроизводимость результатов; удобство качества/аудита	Высокая скорость аналитики; устойчивость и воспроизводимость результатов; удобство качества/аудита	Требуются тяжёлые аналитические запросы и историчность, допустима задержка обновления	DWH/Lakehouse + семантическая витрина; периодические загрузки; индексирование/кэш

Знаниевые подходы: онтологии как «единый язык» семантической интеграции. Онтологический подход основан на том, что корректная интеграция гетерогенных источников невозможна без предварительного согласования понятийной модели предметной области. Если на уровне синтаксиса данные можно привести к общим форматам (таблицы, JSON, события), то на уровне смысла остаются различия в терминологии, классификаторах, единицах измерения и контекстах. Онтология задаёт формальное описание классов, свойств и отношений домена, а также ограничения, которые фиксируют “как правильно понимать данные”. В результате источники получают согласованный словарь, который снижает риск ложных сопоставлений и дублирования сущностей.

В типовом сценарии онтологически-ориентированная интеграция включает следующие шаги:

- 1) *выбор или разработка базовой онтологии*, покрывающей ключевые сущности и отношения всех источников (например, *Sensor, Event, Location, Organization*);
- 2) *семантическая аннотация источников*, при которой элементы схемы источника (поля, сущности, сообщения) сопоставляются понятиям онтологии;
- 3) *приведение данных к семантически совместимому представлению* – либо материализация фактов в форме RDF-триплетов («субъект–предикат–объект»), либо формирование экземпляров классов онтологии;
- 4) *хранение и обработка* интегрированных данных в средствах, поддерживающих семантические запросы (например, SPARQL [14]) и, при необходимости, логический вывод.

Ключевая практическая ценность онтологии в СППР состоит в том, что правила интеграции и запросы формулируются в терминах предметной области, а не в терминах отдельных источников. Например, в сценарии умного города отдельные потоки могут представлять данные о трафике, погоде и инцидентах в разных форматах и с разными атрибутами. При наличии онтологии, где “инцидент” связан с “локацией”, а “погодные условия” – с той же локацией и временем наблюдения, становится возможным единообразно задавать запросы и правила вида: “обнаружить аварии рядом с перегруженными дорогами при температуре ниже 0°C”, не привязываясь к структурам конкретных API или БД.

Важно разграничивать онтологию как схему смысла и данные как факты/наблюдения: онтология фиксирует классы и отношения, а факты (экземпляры) постоянно поступают из потоков и пополняют интегрированное представление. Это разграничение удобно визуализировать (рис. 4): слева располагается слой знаний (классы и свойства), справа – слой экземпляров (конкретные датчики, события, измерения), связанный с онтологией через типизацию и свойства.

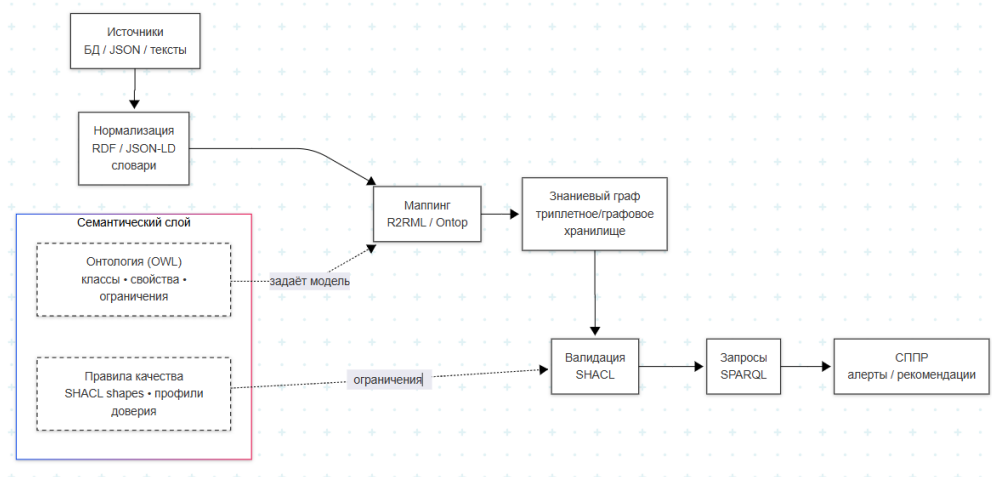


Рис. 4. Онтология (классы/свойства) и экземпляры данных (факты) в семантической интеграции

Ограничениями онтологического подхода являются трудоёмкость разработки и поддержки доменной модели, необходимость сопровождения маппингов источников и потенциальные проблемы масштабируемости при росте объема фактов. На практике эти ограничения снижаются:

- 1) модульным построением онтологии (ядро + доменные расширения);
- 2) автоматизацией части аннотации и сопоставлений;
- 3) сочетанием материализации и виртуализации данных;
- 4) использованием гибридной архитектуры, где онтология задаёт смысловой слой,

а потоковые и федеративные механизмы обеспечивают производительность и актуальность данных.

Знаниевые графы и разрешение сущностей. Знаниевые графы (Knowledge Graphs, KG) развивают онтологический подход, делая акцент не только на формализации понятий, но и на сквозной идентификации объектов реального мира и явном представлении связей между ними. В KG узлы соответствуют сущностям (объектам, событиям, организациям, людям), а ребра – отношениям. Ключевая практическая роль KG в семантической интеграции заключается в том, что разные источники могут описывать одну и ту же сущность фрагментарно и разными словами, а граф позволяет собрать эти описания в единый “профиль” сущности, сохранив происхождение фактов и контекст.

Центральным механизмом при построении KG является разрешение сущностей (entity resolution) и связывание сущностей (entity linking). Под entity resolution понимают задачу выявления, относятся ли записи из разных источников к одному и тому же объекту (например, “ул. Ленина, 10” в реестре и “Lenina 10” в сообщении гражданина). Под entity linking – привязку упоминаний из неструктурированного текста к конкретным узлам графа (например, “ДТП на улице X” → узел *Accident_123* с геокоординатами). На практике эти задачи решаются комбинацией методов:

- 1) правила и нормализация (приведение адресов/имен/единиц к каноническому виду);
- 2) метрики сходства строк и атрибутов;
- 3) вероятностные модели и классификаторы соответствий;
- 4) эмбединги и трансформеры для сравнения текстовых описаний;
- 5) использование контекста графа (соседние связи повышают уверенность в совпадении).

Результат работы entity resolution – операция “слияния узлов” (merge): если два узла признаны описаниями одной сущности, они объединяются, а атрибуты агрегируются с учетом доверия к источникам и правил качества. Важно, что KG обычно хранит не только итоговое значение, но и происхождение: из какого источника пришел факт, когда он получен и какова его надежность. Это повышает объяснимость СППР: система может показать, какие источники подтвердили событие и какие признаки повлияли на решение. Типовой пример для умного города: сообщение в соцсетях “авария на улице X” связывается с узлом события, который также получает подтверждение от дорожного датчика (аномальное падение скорости) и от камеры (распознавание инцидента). После слияния СППР получает единое событие с геопривязкой, временной меткой, уровнем доверия и ссылками на источники.

Рис. 5 иллюстрирует принцип “слияния узлов”: до разрешения сущностей граф содержит два фрагмента описания события из разных источников, после – один объединенный узел с агрегированными атрибутами и указанием происхождения фактов.

Согласование знаний и маппинги источников. Даже при наличии онтологии задача семантической интеграции не решается автоматически: реальные информационные системы часто используют разные словари, классификаторы и модели данных, сформированные независимо друг от друга. В результате возникает необходимость согласования онтологий (ontology alignment) и построения маппингов между локальными схемами источников и глобальной (целевой) моделью. С практической точки зрения, alignment отвечает на вопрос “какие понятия и свойства разных систем означают одно и то же”, а маппинг – “как именно преобразовать данные источника, чтобы они соответствовали целевым понятиям”.

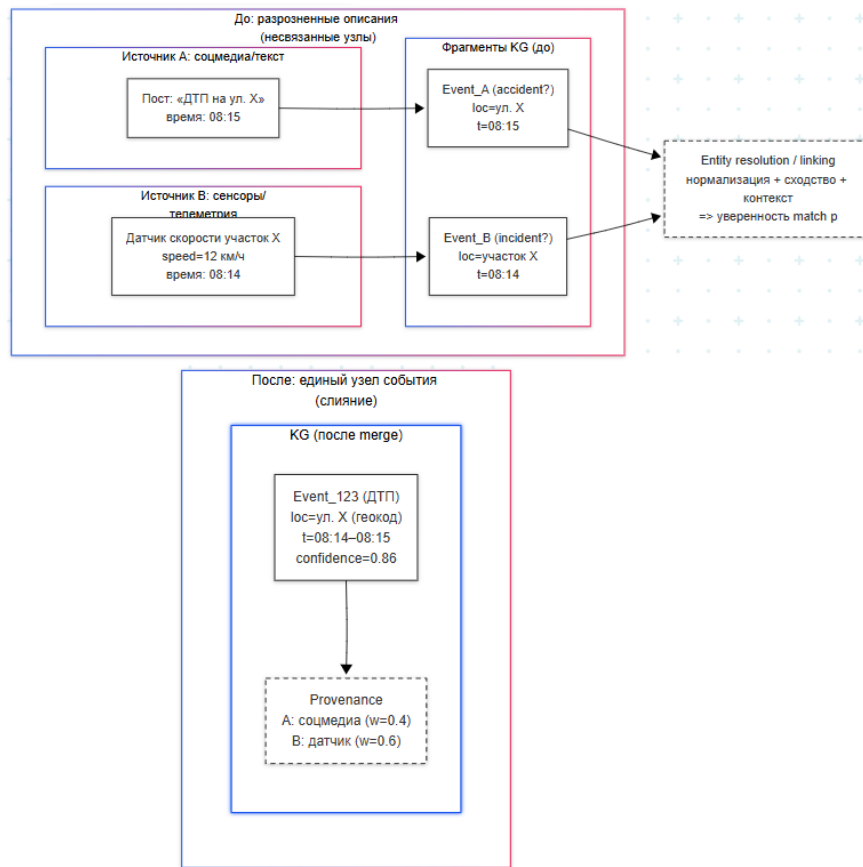


Рис. 5. Разрешение сущностей и «слияние узлов» в знаниевом графе

Методы ontology alignment могут быть автоматическими и полуавтоматическими. Наиболее распространены:

- ◆ лексические и структурные эвристики (сходство названий, синонимы, близость по структуре графа/иерархии);
- ◆ статистические методы (сходство распределений значений, совпадение доменов/типов полей);
- ◆ машинное обучение (включая эмбединги и трансформеры для сопоставления описаний и метаданных);
- ◆ экспертная верификация, необходимая в критичных доменах и при высокой цене ошибки.

В литературе встречается архитектурный шаблон “глобальная онтология + локальные онтологии источников + слой соответствий”. Например, в рамках подхода DIF (Data Integration Framework) формируется глобальная онтология, для каждого источника используется локальная модель, а механизм маппинга соединяет локальные и глобальные элементы, позволяя консолидировать данные [3]. Такой подход снижает связность системы: изменение одного источника требует обновления только его маппингов, а не всей интеграционной схемы.

Для структурированных источников (прежде всего реляционных БД) ключевым инструментом являются маппинги “RDB→RDF”, которые формально описывают, как таблицы, строки и столбцы отображаются в RDF-триплеты и классы/свойства онтологии. На практике широко применяется спецификация R2RML [15], а также системы, которые используют совместимые принципы отображения.

При проектировании интеграции важно выбрать стратегию работы с данными: материализацию либо виртуализацию (рис. 6).

Материализация предполагает извлечение (ETL/ELT) и преобразование данных в RDF/граф (или в унифицированное хранилище) с последующей загрузкой в триплетное/графовое хранилище. Плюсы – быстрые семантические запросы, удобство контроля качества и историчности; минусы – стоимость сопровождения загрузок и риск устаревания копий.

Виртуализация предполагает, что данные остаются в исходных системах, а маппинг (R2RML/Ontop-подход) позволяет выполнять запросы к “виртуальному RDF”: поступающий SPARQL запрос транслируется в SQL/подзапросы и выполняется в источнике, после чего результаты собираются и приводятся к семантическому представлению. Плюсы – актуальность данных и отсутствие копирования; минусы – более сложная оптимизация и зависимость латентности от источников.

Таким образом, ontology alignment и формальные маппинги выступают “мостом” между разнородными системами и семантической моделью. Они обеспечивают масштабируемость внедрения: добавление новых источников и изменение существующих сводится к обновлению соответствий и маппингов, а не к перепроектированию всей СППР.

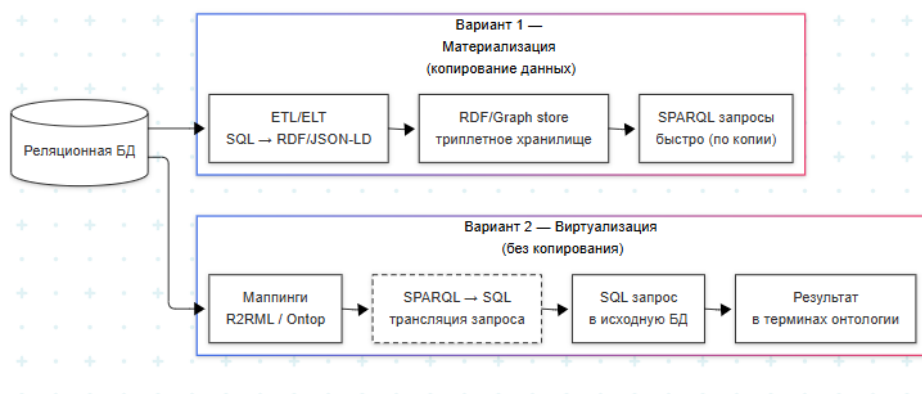


Рис. 6. Два сценария интеграции: материализация данных и виртуализация через маппинги

Семантическое обогащение неструктурированных потоков (NLP/CV). Для СППР критически важны не только структурированные источники (БД, реестры, телеметрия), но и неструктурированные потоки: текстовые сообщения, отчёты, обращения граждан, новости и социальные медиа, а также изображения и видео. Именно такие источники часто дают ранние сигналы о событиях (инцидентах, сбоях, изменениях ситуации), однако в “сыром” виде они плохо пригодны для автоматической интеграции: содержат неоднозначные упоминания, шум, жаргон, неполные описания и контекстные ссылки. Поэтому в современных системах интеграции применяется семантическое обогащение (semantic enrichment) – преобразование неструктурированного контента в формализованные факты, согласованные с онтологией и знаниевым графом (рис. 5).

Семантическое обогащение в текстовых потоках обычно включает следующие этапы:

- 1) *предобработка* (язык, токенизация, нормализация, удаление спама/дубликатов);
- 2) *извлечение сущностей (NER)* – выделение объектов (персон, организаций, локаций, событий, показателей);
- 3) *извлечение отношений (RE)* – выявление связей между сущностями (например, “инцидент произошёл в локации”, “организация заключила контракт”, “датчик измерил параметр”);
- 4) *нормализация и канонизация* – приведение сущностей к стандартному виду (адреса, единицы измерения, временные выражения);

5) *entity linking* – привязка упоминаний к конкретным узлам графа знаний (или создание новых узлов при отсутствии);

6) *формирование фактов* в виде триплетов/графовых утверждений с метаданными (время, источник, доверие).

Для мультимодальных потоков (изображения/видео) аналогичную роль выполняют методы компьютерного зрения: детекция объектов и сцен (например, “ДТП/скопление людей/затопление”), распознавание текстов на изображении (OCR) и последующая привязка результата к сущностям графа (локация, время, тип происшествия). Важный практический момент – хранение provenance: любая извлечённая сущность и связь должны сопровождаться указанием источника, времени извлечения и оценкой уверенности модели. Это позволяет СППР управлять качеством: понижать вес сомнительных наблюдений, подтверждать события по нескольким независимым каналам и объяснять пользователю, на каких данных основана рекомендация.

Семантическое обогащение существенно расширяет область применения онтологий и знаниевых графов: они перестают быть “слоем для БД” и начинают работать как единый контур интеграции, включающий тексты и мультимедиа. В результате СППР получает не просто поток сообщений, а поток структурированных событий и фактов, которые можно объединять с телеметрией и реестрами, валидировать правилами качества и использовать для оперативных решений.

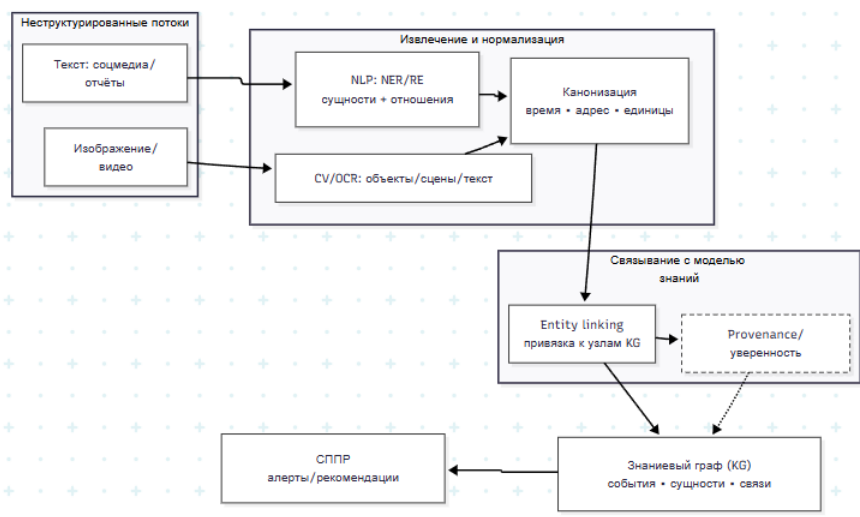


Рис. 7. Семантическое обогащение неструктурированных потоков:
NLP/CV → linking → KG

Алгоритмические подходы: data fusion и управление неопределённостью. Даже при наличии единой семантической модели (онтологии/графа знаний) интеграция в СППР требует алгоритмических решений, которые отвечают на вопрос: как объединять наблюдения из разных источников, если они поступают с разной частотой, имеют разные погрешности и могут противоречить друг другу. Эта группа методов относится к классу data fusion (слияние данных) и включает агрегирование измерений, обновление гипотез о событиях, оценку состояния системы, а также детектирование инцидентов в режиме реального времени. В мульти-источниковых сценариях цель fusion – получить более точную и устойчивую оценку, чем даёт любой отдельный источник, при этом сохранив информацию о доверии и происхождении фактов.

Неопределённость в потоке данных проявляется в нескольких формах:

- 1) шум и случайные ошибки измерений (типично для сенсоров);
- 2) латентность и пропуски (часть источников обновляется редко или теряет сообщения);

- 3) семантические неоднозначности (одно и то же событие описано разными словами);
- 4) конфликтующие свидетельства (соцмедиа сообщает об аварии, а датчики её не подтверждают).

Поэтому алгоритмы fusion обычно включают:

- 1) нормализацию (приведение к общим единицам);
- 2) взвешивание источников по надежности;
- 3) механизм разрешения конфликтов;
- 4) обновление оценки по мере поступления новых свидетельств.

Классический набор методов, применяемых в data fusion, включает вероятностные и эвристические подходы. Для числовых временных рядов и задач оценивания состояния часто используют *фильтр Калмана* и его расширения, позволяющие сглаживать шум и оценивать скрытые переменные системы. Для событийных сценариев, где поступают неоднородные “свидетельства”, применяются *байесовские методы* (обновление апостериорных вероятностей гипотез) и *теория Демпстера–Шафера*, удобная для объединения частичных и неопределённых свидетельств и явного представления “массы доверия” к гипотезам. В задачах, требующих интерпретируемости при неоднозначных правилах, используются *нечёткие модели* (fuzzy logic), где экспертные знания формализуются через лингвистические правила (“высокая температура” и “резкое падение скорости” равно “вероятность инцидента высокая”).

Для современных СППР важным является *связующее звено* между контуром качества и fusion: результаты профилирования качества и валидации формируют *веса доверия источникам* (или confidence к наблюдениям), которые затем участвуют в объединении. Практически это реализуется как взвешенные агрегаты для числовых данных, ранжирование и пороговые решения для инцидентов, а также гибридные схемы, где разные методы используются последовательно (например, Калман для сенсорной части + Демпстер–Шафер для событийных свидетельств + байесовское обновление итоговой уверенности). Такой подход особенно эффективен в real-time СППР, где нужно быстро реагировать, но при этом снижать риск ложных срабатываний за счёт мульти-источникового подтверждения (рис. 8).

Таблица 3

Методы data fusion и управление неопределённостью в СППР

Метод	Для каких данных	Что даёт	Типовые задачи в СППР
Взвешенная агрегация / voting	однотипные сенсоры, показатели	быстрое объединение с весами доверия	агрегирование телеметрии, первичный скоринг
Фильтр Калмана (EKF/UKF)	временные ряды	оценка состояния, сглаживание, прогноз	трафик, производство, предиктивное обслуживание
Байесовское обновление	гипотезы/события, вероятн. признаки	обновление вероятности события	диагностика, оценка риска/инцидента
Демпстер–Шафер	неоднородные свидетельства	объединение частичных/конфликтных данных	подтверждение инцидентов (сенсоры+текст+CV)
Нечёткая логика	экспертные признаки	интерпретируемые правила “если–то”	триггеры реакции, риск-скоринг
ML-модели (ансамбли/NN)	смешанные признаки (в т. ч. NLP/CV)	извлечение сложных паттернов	классификация событий, прогноз, приоритизация
KG-контекст (graph-aware)	факты/связи в KG	учет контекста и связности	консолидация событий, объяснение решений

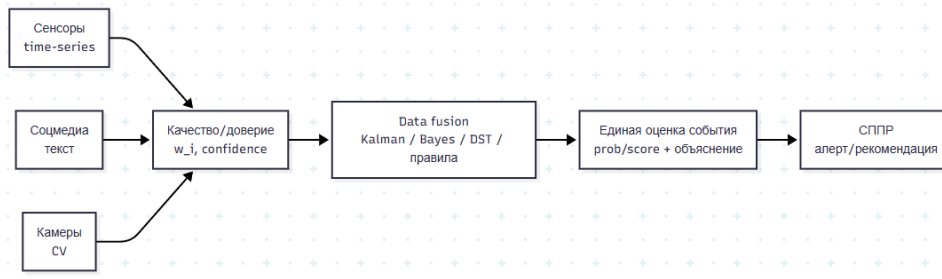


Рис. 8. “Fusion-узел”: как доверие и потоки сходятся в оценку события

Контроль качества и семантическая валидация (SHACL/правила). Контроль качества данных является обязательным компонентом семантической интеграции: даже идеально согласованная онтология не компенсирует шум, пропуски, задержки и противоречия, характерные для многопоточных источников. В СППР качество влияет непосредственно на надежность рекомендаций, поэтому контур качества обычно строится как непрерывный процесс, работающий как на входе (до попадания фактов в интегрированное представление), так и на уровне уже сформированного графа знаний.

Практически контроль качества включает четыре взаимосвязанных слоя.

На слое *профилирования и мониторинга качества* происходит измерение полноты, частоты обновления, задержки, доли пропусков и аномалий для каждого источника и типа данных. Эти метрики используются для расчёта веса доверия источнику и выбора политики обработки (например, “для источника X применять более строгие фильтры и подтверждение по второму каналу”).

На слое *очистки и фильтрации* осуществляется дедупликация, подавление выбросов, удаление спама и сервисных сообщений, нормализация единиц измерения и форматов времени/географии. Для потоков важны online-проверки (окна, пороги, rate-limit), позволяющие отбрасывать явные ошибки до того, как они повлияют на решения.

На слое *семантической валидации* реализован контроль соответствия фактов доменной модели (типам, обязательным свойствам, диапазонам, допустимым связям). В отличие от “обычной” валидации схемы, семантическая проверяет корректность в контексте предметной области: например, что событие “ДТП” обязательно имеет время и локацию, а скорость измерения не может быть отрицательной; что датчик, измеряющий скорость, связан с дорожным участком; что временная метка находится в разумном диапазоне относительно текущего времени и т.п.

Последний слой – *управление неопределенностью и конфликтами*. При обнаружении противоречий между источниками система может:

- 1) понижать уверенность/вес спорных фактов;
- 2) переводить событие в статус “требуется подтверждения”;
- 3) инициировать повторную проверку или запрос к дополнительному источнику;
- 4) сохранять несколько альтернативных значений с указанием происхождения и уверенности.

Для RDF/knowledge graph-представлений широко применяется язык ограничений SHACL (Shapes Constraint Language) [16], позволяющий описывать правила валидации как “шаблоны” (shapes), которые должны выполняться для узлов определенного типа. SHACL удобен тем, что делает правила качества явными и проверяемыми, а также позволяет унифицировать валидацию для данных, поступающих из разных источников. В потоковых СППР SHACL-валидация используется либо на этапе пакетной проверки графа, либо выборочно “на входе” для критичных сущностей (событий и ключевых измерений), чтобы снизить риск ложных срабатываний.

На рис. 9 приведен пример SHACL-ограничений для события типа *Accident*: обязательны локация и время наблюдения, требуется хотя бы одно свидетельство (*evidence*), а также задается допустимый диапазон уверенности события. Дополнительно введено доменное правило: временная метка не должна существенно уходить в будущее (допускается небольшой “люфт” из-за задержек доставки и различий системного времени).

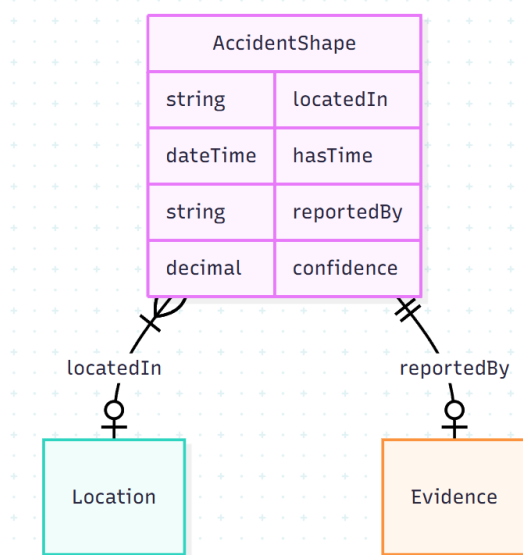


Рис. 9. Пример SHACL-ограничений (NodeShape) для события *Accident*: обязательные свойства и связи с классами *Location* и *Evidence*

Практическая ценность подобных shapes заключается в том, что проверка качества становится формализованной и воспроизводимой: правила можно версионировать, применять к разным источникам и запускать как в пакетном режиме (аудит графа), так и в потоковом режиме на критичных сущностях. При обнаружении нарушений система, как правило, не ограничивается “ошибкой”, а применяет политику обработки (quality policy). На практике используются три базовые стратегии:

- 1) *reject* – отклонение факта/события (например, отрицательная скорость или координаты вне диапазона);
- 2) *quarantine* – помещение в карантин/очередь на уточнение (например, событие без локации или времени);
- 3) *soft-fail* – пропуск с понижением доверия (например, неполные данные из источника с низкой надежностью), когда важно не терять сигнал, но необходимо отразить его неопределенность.

Выбор стратегии зависит от критичности атрибута и домена СППР: в оперативных сценариях часто допустим *soft-fail* для раннего предупреждения, но с обязательным последующим подтверждением (кросс-проверкой) и обновлением *confidence*. Таким образом, контур качества выполняет две ключевые функции:

- 1) защищает СППР от “плохих” данных и снижает риск ошибочных рекомендаций;
- 2) формирует метрики доверия и уверенности, которые далее используются в алгоритмах слияния и принятия решений.

Визуально взаимодействие этапов контроля качества удобно представить как замкнутый контур “профиль качества → фильтрация → семантическая валидация → веса доверия → fusion/решение”.

Архитектурно-гибридные решения: потоковая обработка, федерация и семантический слой. Практическая реализация семантической интеграции в СППР почти всегда требует архитектурно-гибридного подхода: семантический слой (онтологии, граф

знаний, правила валидации) должен быть встроен в инфраструктуру сбора и обработки данных, способную работать с высокой скоростью поступления событий и большим числом источников. Поэтому современные решения сочетают:

- 1) потоковую обработку для низкой задержки;
- 2) семантическое моделирование и KG для согласования смысла и сквозной идентификации сущностей;
- 3) контур качества/валидации для надежности результатов;
- 4) федеративный доступ для случаев, когда данные нельзя или нецелесообразно копировать в единое хранилище.

В потоковом слое выполняются операции, которые критичны по времени:

- ◆ прием событий;
- ◆ оконные агрегации;
- ◆ корреляция;
- ◆ первичный детект инцидентов;
- ◆ расчёт признаков.

На этом же этапе обычно производится нормализация форматов (например, приведение JSON-событий к каноническим полям), а также первичная фильтрация и присвоение меток качества. Типовой стек включает брокер сообщений и потоковый движок, позволяющий обрабатывать события по “event-time”, выдерживать задержки и переупорядочивание сообщений.

Семантический слой может быть реализован двумя способами:

- 1) материализованно, когда факты загружаются в KG/триплетное хранилище и далее доступны для SPARQL-запросов и вывода;
- 2) виртуально, когда KG выступает как “семантическая витрина” поверх источников, а запросы транслируются к исходным системам через маппинги и федерацию.

На практике часто применяется компромисс: материализуются наиболее важные, “оперативные” сущности (ключевые события, справочники, идентификаторы), а тяжелые детальные данные остаются в источниках и подключаются по мере необходимости.

Федеративный доступ становится критичным в двух ситуациях:

- 1) данные живут в разных хранилищах (SQL/NoSQL/объектные хранилища/API) и их копирование дорого или запрещено;
- 2) требуется максимально актуальный ответ в момент запроса.

В этом случае СППР использует единый слой запросов, который распределяет подзапросы по источникам и собирает результат, а семантические описания источников (метаданные, маппинги, словари) помогают корректно “сшивать” данные. Ограничением федерации является рост латентности и сложность оптимизации — поэтому в архитектуре обычно применяются кэширование, материализация горячих данных и предвычисление признаков.

В итоге архитектурно-гибридная схема позволяет построить СППР, которая одно-временно реагирует в реальном времени, сохраняет семантическую согласованность данных, умеет контролировать качество и доверие, и не требует обязательного полного копирования всех данных в единое хранилище. Обобщённая конфигурация таких решений показана на рис. 10.

Типовой конвейер семантической интеграции потоков и соответствие технологий. Рассмотренные выше методы (онтологии и знаниевые графы, entity linking, семантическое обогащение, data fusion, контроль качества и архитектурные решения) на практике реализуются не в виде одного “универсального” модуля, а как конвейер взаимосвязанных этапов, через которые проходят данные от момента поступления из источников до формирования рекомендаций СППР. Такой конвейер позволяет:

- 1) разделить ответственность компонентов;
- 2) масштабировать систему по узким местам (например, отдельно масштабировать обработку потоков или извлечение сущностей из текстов);
- 3) внедрять контроль качества как непрерывный процесс;
- 4) обеспечить трассируемость фактов и их происхождения.

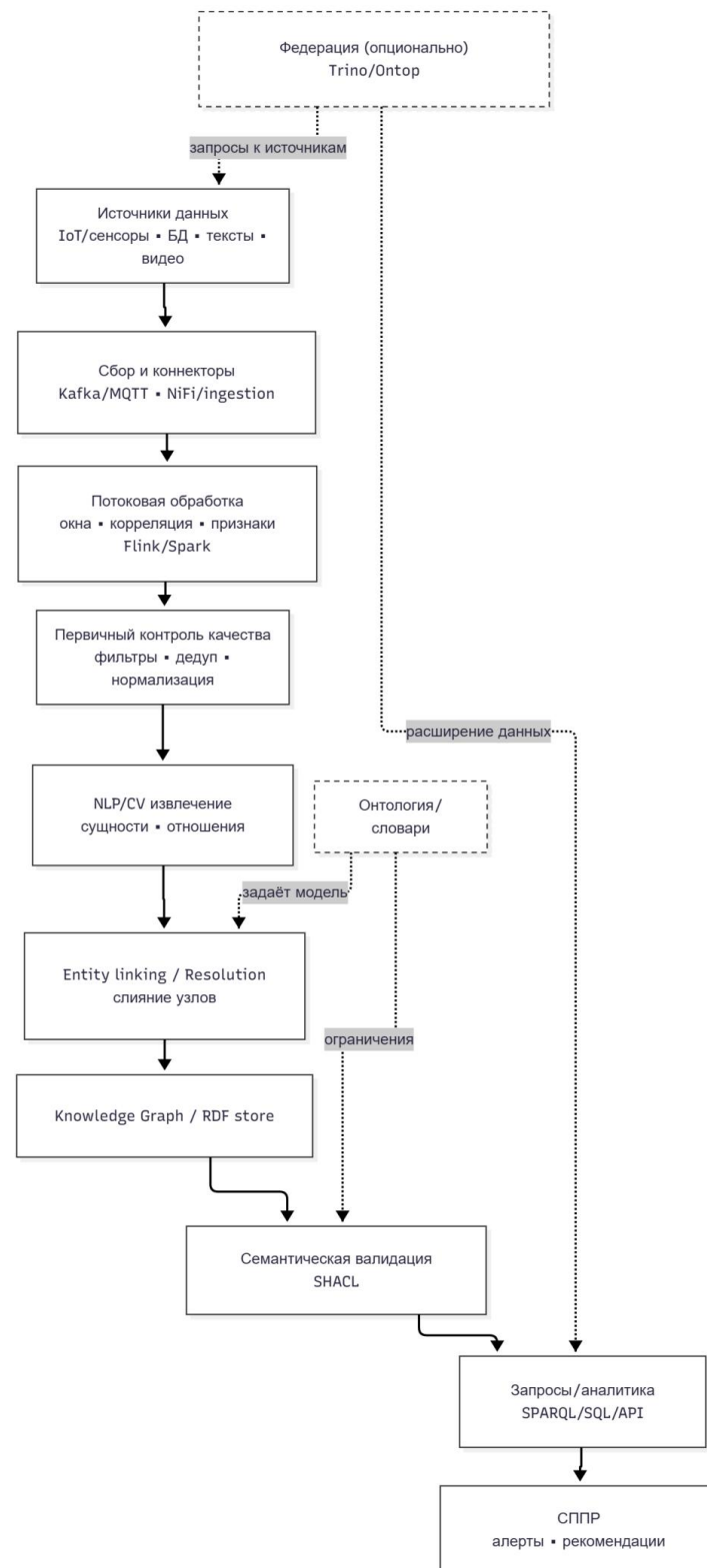


Рис. 10. Архитектура семантической интеграции для СПДР: потоковый слой + KG/валидация + федерация источников

Типовая последовательность шагов включает:

- 1) сбор данных (стрим/файлы/запросы);
- 2) потоковую обработку (окна, корреляция, детект событий);
- 3) нормализацию/гармонизацию форматов и единиц;
- 4) семантическое моделирование домена;
- 5) маппинг структурированных источников (включая RDB→RDF);
- 6) валидацию качества (в том числе через SHACL);
- 7) семантические запросы к интегрированному представлению.

На каждом этапе могут применяться как знаниевые, так и алгоритмические методы: например, извлечение сущностей и связывание с KG выполняется на этапе обогащения, а веса доверия и слияние данных – на стыке контроля качества и принятия решения.

Сопоставление этапов интеграционного конвейера с типовыми стандартами и примерами технологий приведено в табл. 2. Эта таблица может использоваться как практическая “карта” при проектировании СППР: она помогает определить, какие компоненты являются обязательными для заданного класса задач (например, real-time мониторинг), а какие – опциональными и зависят от ограничений (например, федерация при запрете на копирование данных).

Таким образом, семантическая интеграция в СППР представляет собой комбинацию семантического слоя (онтологии/KG) и инженерной инфраструктуры (потоки, коннекторы, федерация), дополненную алгоритмами слияния и контуром качества. В следующем разделе рассматриваются примеры внедрения этих подходов в прикладных доменах и обсуждаются достигнутые эффекты (точность, скорость реакции, устойчивость к неоднозначностям и качеству данных).

Практические эффекты применения семантической интеграции в СППР. Эффективность семантической интеграции в СППР оценивается не только по тому, насколько корректно объединяются данные из разных источников. Важно также, как меняются прикладные показатели работы системы. Обычно смотрят, стала ли ситуационная картина более полной, повысилась ли точность рекомендаций и уменьшилось ли время реакции. Отдельное внимание уделяют устойчивости к шуму и противоречиям между источниками. Анализ публикаций и практических внедрений показывает, что наиболее заметный эффект появляется там, где система опирается на несколько каналов с разной надежностью и скоростью обновления. Существенную роль играет наличие неструктурированной информации, которая поступает в виде текстов и мультимедиа и требует извлечения смысловых фактов. Кроме того, во многих прикладных областях важно, чтобы результаты были объяснимыми и трассируемыми. В этом разделе обобщаются прикладные эффекты внедрения по типовым доменам и рассматриваются факторы, которые сильнее всего влияют на результат. К ним относятся выбранная архитектура, качество маппингов между источниками и моделью знаний, зрелость контура валидации, а также правила учета доверия к источникам.

Повышение полноты и точности ситуационной осведомлённости. Одним из ключевых эффектов семантической интеграции в СППР становится рост ситуационной осведомлённости. Под этим понимается способность системы формировать целостное и непротиворечивое представление о происходящем на основе данных из разных источников. Полнота здесь означает, что в интегрированном представлении присутствуют все значимые объекты и события, а также их важные атрибуты. Точность означает, что эти объекты корректно интерпретированы, не задублированы и правильно связаны друг с другом. Например, событие должно быть привязано к корректной локации и времени, а показатели – к нужным сущностям и единицам измерения.

Рост полноты достигается прежде всего за счёт расширения охвата источников. Структурированные системы обычно дают устойчивые данные, но описывают ситуацию “частично”. Текстовые потоки и сообщения людей часто дают ранние сигналы, а также контекст, который отсутствует в телеметрии. Когда применяется семантическое обогащение на основе NLP и компьютерного зрения, неструктурированный контент преобразуется в формализованные факты. После этого такие факты можно сопостав-

лять с показаниями датчиков и записями реестров. В результате система реже пропускает важные события только потому, что они появились в канале, который изначально не был “табличным”.

Повышение точности обеспечивается несколькими механизмами, которые работают совместно, но выполняют разные задачи. Онтологии и знаниевые графы задают общую смысловую модель. За счёт этого уменьшается количество ошибок, связанных с разной терминологией и разными правилами интерпретации. Механизмы entity resolution и entity linking снижают дублирование. Они позволяют “сшивать” записи из разных источников, если речь идёт об одном и том же объекте или событии. При этом система может сохранять происхождение фактов и уверенность в них, что важно для объяснимости.

Отдельную роль играет контур качества. Он отсекает явно некорректные данные и снижает влияние сомнительных сообщений. Семантическая валидация помогает проверять факты в контексте предметной области. Она выявляет ситуации, когда событию не хватает обязательных атрибутов, когда значения физически невозможны или когда связи между сущностями противоречат модели. Если данные всё же потенциально полезны, система может не удалять их полностью. Вместо этого она может снизить доверие и запросить подтверждение по другим каналам. Алгоритмы слияния данных дополняют этот контур. Они позволяют повышать уверенность события, когда разные независимые источники дают согласующиеся признаки.

В прикладных кейсах это приводит к тому, что СППР формирует не набор разрозненных сигналов, а единую “карту событий”. Для каждого события становится понятнее, что произошло, где и когда это случилось, какими источниками это подтверждается и насколько этому можно доверять. В городских сценариях интеграция сообщений граждан с дорожной телеметрией помогает быстрее обнаруживать инциденты и одновременно снижать число ложных тревог за счёт подтверждения по нескольким каналам. В здравоохранении графовые модели упрощают согласование терминов и объединение клинических данных с внешними знаниями, что снижает неоднозначность и повышает связность фактов. В бизнес-сценариях объединение разнородных потоков повышает точность прогнозов и рекомендаций по сравнению с решениями, построенными на одном источнике [4–6].

Чтобы измерять эффект полноты и точности, в исследованиях и внедрениях обычно фиксируют несколько показателей. В первую очередь оценивают долю найденных событий относительно эталона и точность обнаружения. Затем измеряют качество “сшивки” сущностей между источниками. Дополнительно считают количество дублей до и после entity resolution. Часто отдельно проверяют долю фактов, прошедших правила качества и семантическую валидацию. Эти метрики позволяют показать, что улучшилось не только количество данных, но и их пригодность для принятия решений.

Сокращение времени реакции и автоматизация решений. Второй важный практический эффект семантической интеграции связан со скоростью реакции СППР. В задачах оперативного управления ценность данных снижается по мере роста задержки: ситуация может измениться быстрее, чем оператор успеет собрать информацию из разных систем и согласовать её смысл. Семантическая интеграция уменьшает задержку за счёт того, что данные заранее проходят единый конвейер нормализации, связывания сущностей и проверки качества. В результате система быстрее формирует понятное событие и выдаёт рекомендацию, а не поток разрозненных сигналов.

Сокращение времени реакции обычно обеспечивается сочетанием трёх механизмов (рис. 11):

1. *Потоковая обработка и корреляция.* События обрабатываются “на лету”, агрегируются по окнам времени и связываются по пространственному и временному контексту. Это сокращает время на ручную “сборку” картины из фрагментов.
2. *Семантическое связывание и устранение дублей.* Entity linking/resolution позволяет быстро объединять признаки одного события из разных источников в единый объект (инцидент/событие), что уменьшает задержки на сопоставление и снижает число ложных уведомлений.

3. *Контур качества и подтверждение по нескольким каналам.* SHACL/правила качества отсекают явно некорректные факты, а механизмы доверия к источникам позволяют быстро повышать уверенность события при наличии независимых подтверждений. Это ускоряет принятие решения без роста ложных тревог.

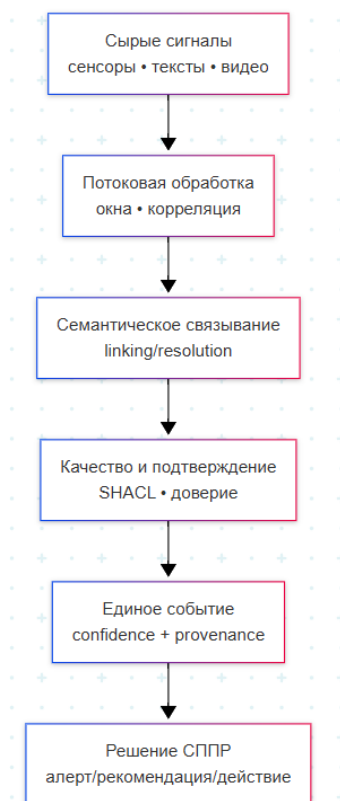


Рис. 11. Конвейер сокращения времени реакции в СППР: от сырого сигнала к подтверждённому событию и действию

Автоматизация решений в СППР обычно развивается поэтапно и зависит от домена и требований к рискам:

1. *Поддержка оператора:* система формирует алерт и объяснение, а решение принимает человек.
2. *Полуавтоматический режим:* часть действий выполняется автоматически, но требует подтверждения или может быть отменена.
3. *Автоматический запуск процедур:* система выполняет заранее определённые действия, если выполнены требования качества и уверенность события достаточна.

Чтобы ускорение не приводило к росту ошибок, в реальных внедрениях вводят политику обработки событий при нарушении качества:

- 1) *reject* – отклонение факта или события (физически невозможные значения, критические ошибки);
- 2) *quarantine* – карантин до уточнения (не хватает обязательных атрибутов);
- 3) *soft-fail* – пропуск с понижением уверенности (сохранение слабого сигнала с пометкой доверия).

Эффект сокращения времени реакции измеряют несколькими метриками:

- 1) задержка от поступления первичного сигнала до формирования события в интегрированном представлении;
- 2) время до выдачи рекомендации оператору;

- 3) в автоматизированных сценариях – время полного цикла от события до действия;
- 4) доля ложных тревог и доля событий, подтвержденных несколькими источниками.

Эффективность и экономический эффект. Помимо роста точности и сокращения времени реакции, семантическая интеграция даёт прикладной эффект в виде повышения эффективности процессов и снижения затрат. Это связано с тем, что решения начинают приниматься на более полной и согласованной картине. Система быстрее выявляет отклонения и причины, а также точнее определяет приоритеты действий. В результате уменьшаются простои, снижаются потери ресурсов и улучшаются целевые показатели работы организаций и городских служб [17].

Наиболее наглядно экономический эффект проявляется в сценариях, где есть измеримые операционные показатели. В промышленности это связано с предиктивным обслуживанием и управлением производством. Интеграция телеметрии оборудования, журналов ремонтов и данных ERP позволяет точнее определять состояние активов и планировать обслуживание до наступления отказа. При этом знаниевый граф помогает связать симптомы с конкретными узлами оборудования и историями неисправностей, а контур качества снижает риск ошибочных срабатываний из-за шумных датчиков. В городских системах эффект выражается в более точном распределении ресурсов. Это касается транспорта, коммунальной инфраструктуры и реагирования на инциденты. Когда события подтверждаются несколькими источниками, диспетчерские службы тратят меньше времени на проверку и меньше ресурсов на ложные выезды.

Экономическая результативность усиливается, когда система не просто фиксирует события, а поддерживает полный цикл управления. Сначала происходит обнаружение и подтверждение инцидента. Затем система предлагает оптимальное действие или автоматически запускает процедуру. После этого результаты фиксируются и используются для уточнения правил и веса доверия к источникам. Такая обратная связь постепенно снижает количество лишних действий и повышает точность планирования (рис. 12).



Рис. 12. Цикл повышения эффективности

Эффект внедрения можно представить через набор измеримых ключевых показателей эффективности (KPI - Key Performance Indicators). Чаще всего используют показатели, связанные с простоем и затратами, а также с качеством сервиса. Для промышленности это длительность простоев и расходы на обслуживание. Для городских сценариев это

время реагирования, нагрузка на службы и доля подтверждённых инцидентов. Для бизнес-сценариев это точность прогнозов и эффективность решений, влияющих на выручку и потери (табл. 4).

Таблица 4

Примеры KPI для оценки экономического эффекта семантической интеграции

Домен	KPI “до/после”	Как связан с интеграцией
Промышленность	простои, MTTR, затраты на ТО	точнее диагностика и планирование по мульти-источниковым данным
Умный город [18]	время реагирования, ложные выезды	подтверждение событий по нескольким каналам + качество данных
Бизнес/туризм	точность прогноза спроса, скорость пересчёта	fusion + потоковые признаки + обогащение текстов [19]
Здравоохранение	согласованность данных, время поиска контекста	KG объединяет ЭМК и знания, снижает ручные проверки

Улучшение качества данных и накопление знаний. Помимо прямых эффектов в виде ускорения реакции и повышения точности, семантическая интеграция даёт важный “вторичный” результат. Когда данные из разных систем приводятся к общей модели и связываются на уровне сущностей, становятся заметны ошибки, пробелы и противоречия, которые раньше оставались скрытыми внутри отдельных источников. Это приводит к росту качества данных в целом и формированию устойчивой базы знаний, которая может использоваться повторно в разных задачах СППР (табл. 5).

Таблица 5

Типовые проблемы качества и результат семантической интеграции

Проблема	Как выявляется	Что меняется после интеграции
Дубли сущностей	entity resolution	единые идентификаторы, меньше дублей
Пропуски атрибутов	SHACL minCount	карантин/дозаполнение
Некорректные значения	диапазоны/правила	отбрасывание/понижение доверия
Конфликты источников	provenance + веса	подтверждение по нескольким каналам

Улучшение качества происходит за счёт нескольких механизмов.

1. *Выявление несоответствий между источниками.* При сведении данных в единый граф знаний обнаруживаются расхождения в атрибутах, датах, единицах измерения и классификаторах. Например, один реестр может содержать устаревшую информацию, а фактические данные сенсоров или отчётов показывают иное состояние объекта.

2. *Дедупликация и уточнение идентичностей.* Entity resolution снижает количество дублей и “размноженных” сущностей. В результате справочники и реестры становятся более консистентными, а отчёты – более сопоставимыми.

3. *Формализация правил корректности.* SHACL и доменные правила переводят требования к данным из неявных “ожиданий” в проверяемые ограничения. Со временем это повышает дисциплину формирования данных в источниках, потому что ошибки начинают фиксироваться и возвращаться ответственным системам.

4. *Обратная связь и исправление источников.* Практика внедрений показывает, что интеграционная платформа становится не только “потребителем”, но и инструментом контроля качества. Нарушения правил качества могут автоматически формировать отчёты для владельцев источников или запускать процедуры корректировки.

Накопление знаний проявляется в том, что граф знаний постепенно превращается в “память” системы. Он хранит не только текущие состояния объектов, но и связи между ними, историю событий, подтверждения и происхождение фактов. Это облегчает объяснимость решений: система может показать, какие факты и из каких источников привели к рекомендации. Кроме того, знания становятся переносимыми между задачами. Одна и та же семантическая модель может использоваться для мониторинга, анализа причин, прогнозирования и планирования действий.

Практически это означает, что организация получает единое семантически согласованное пространство данных, которое со временем становится более чистым и более полезным. Даже если первоначальная цель внедрения была связана с оперативным реагированием, побочным результатом часто становится рост качества мастер-данных, повышение доверия к аналитике и снижение затрат на ручную сверку информации.

Ограничения и риски внедрения. Несмотря на выраженные преимущества, внедрение семантической интеграции в СППР связано с рядом ограничений и рисков, которые необходимо учитывать при проектировании и эксплуатации. В отличие от “классической” интеграции на уровне форматов и ETL, семантический подход затрагивает смысловые модели, правила качества и процессы принятия решений. Поэтому ошибки на этапе моделирования или эксплуатации могут приводить не только к техническим сбоям, но и к неправильным управленческим выводам.

Ключевые риски можно сгруппировать следующим образом.

1. *Трудоёмкость разработки и сопровождения онтологий и маппингов.* Построение доменной модели, согласование терминов и настройка отображений источников на общую схему требуют времени и участия экспертов. При расширении числа источников объём работ растёт, а качество маппингов становится критичным фактором для точности интеграции. Дополнительная сложность возникает при изменениях в источниках, когда схемы “дрейфуют” и маппинги нужно регулярно обновлять.

2. *Масштабируемость и производительность.* Знаниевые графы и семантические запросы дают гибкость и выразительность, но могут становиться узким местом при больших объёмах событий и высокой частоте обновлений. В реальном времени особенно чувствительны операции связывания сущностей, валидации и сложные запросы, которые требуют оптимизации, кэширования и иногда частичной материализации данных.

3. *Риск накопления “шума” в графе знаний.* Если потоковые данные поступают без строгого контроля качества или с низкой надёжностью источников, граф может быстро насыщаться ложными и дублирующимися фактами. Это ухудшает качество рекомендаций и усложняет последующую очистку. Поэтому необходимы политики обработки нарушений качества, а также стратегии жизненного цикла фактов, включая устаревание, архивирование и удаление.

4. *Сложности с конфликтами источников и неопределённостью.* В мульти-источниковой среде противоречия неизбежны. Если веса доверия заданы неверно или подтверждение по нескольким каналам реализовано формально, система может либо “не замечать” реальные события, либо генерировать ложные тревоги. Для снижения риска требуется аккуратная настройка правил слияния, хранение provenance и периодическая калибровка доверия к источникам.

5. *Эксплуатационные риски и дефицит компетенций.* Поддержка семантических компонентов требует специалистов, которые понимают одновременно предметную область, данные, графовые модели и потоковую инженерию. При недостатке компетенций растёт вероятность ошибок конфигурации, деградации качества и “зарастания” системы техническим долгом.

6. *Безопасность, конфиденциальность и соответствие требованиям.* Интеграция объединяет данные из разных контуров, что повышает риск несанкционированного доступа и повторной идентификации персональных данных. Особенно чувствительны медицинские и финансовые сценарии. Помимо технических мер защиты, требуется управление доступом на уровне сущностей и атрибутов, ведение аудита, а также корректная работа с правами и политиками использования данных.

7. *Риск потери прозрачности при использовании ML.* Когда интеграция включает модели машинного обучения для извлечения сущностей и принятия решений, возрастает вероятность скрытых ошибок из-за дрейфа данных и изменения контекста. Это требует мониторинга качества моделей, контроля версий и обязательного сохранения объяснимых следов в графе знаний.

С практической точки зрения минимизация указанных рисков достигается за счёт организационных и технических мер. Важно заранее определить границы доменной модели, выбрать реалистичный уровень детализации онтологии и внедрять интеграцию поэтапно, начиная с наиболее критичных сущностей и сценариев. Требуется формализовать правила качества и политики обработки нарушений, а также обеспечить хранение provenance и механизмов пересмотра доверия к источникам. Для производительности обычно применяется гибридная стратегия: материализация “горячих” данных, кэширование, ограничение тяжёлых запросов в реальном времени и перенос части аналитики в пакетный контур.

Заключение. В работе выполнен обзор технологий семантического анализа и интеграции гетерогенных информационных потоков в системах поддержки принятия решений. На основе анализа публикаций 2018–2025 гг. обобщены подходы к обеспечению семантической совместимости данных и показано, как они встраиваются в потоковые и гибридные архитектуры СППР.

Основные результаты можно сформулировать следующим образом.

1. Показано, что ключевым препятствием для интеграции в СППР остаётся не синтаксическая, а семантическая неоднородность данных. Даже при совпадении форматов различия в терминологии, единицах измерения, классификаторах и контексте приводят к ошибкам интерпретации и дублированию сущностей.

2. Систематизированы методы семантической интеграции и их роль в типовом конвейере: нормализация и гармонизация, семантическое моделирование, маппинг источников, связывание сущностей, контроль качества и семантическая валидация, семантические запросы и федерация. Показано, что максимальная практическая ценность достигается при комбинировании знаниевого слоя (онтологии/граф знаний) с алгоритмами слияния данных и механизмами доверия к источникам.

3. Отмечено, что для неструктурированных потоков решающую роль играют методы NLP и компьютерного зрения, которые обеспечивают извлечение сущностей и отношений и превращают текст/мультимедиа в формализованные факты, пригодные для включения в граф знаний и последующего анализа.

4. На основании рассмотренных кейсов показано, что применение семантической интеграции в СППР повышает полноту и согласованность ситуационной картины, сокращает время реакции за счёт потокового конвейера и снижает долю ложных срабатываний за счёт валидации и подтверждения по нескольким каналам. В ряде прикладных работ фиксируются улучшения точности порядка 15–20% и существенное снижение времени отклика при многоисточниковом объединении данных [5, 20].

5. Выделены ограничения внедрения: трудоёмкость создания и сопровождения онтологий и маппингов, проблемы масштабируемости графов и запросов в реальном времени, риск накопления “шума” при слабом контуре качества, а также требования к безопасности и управлению доступом при интеграции чувствительных данных.

Практическая значимость работы состоит в том, что предложенная структура методов и конвейерная логика интеграции могут служить ориентиром при проектировании СППР для умного города, промышленности и здравоохранения. Перспективы дальнейших исследований связаны с автоматизацией семантического согласования (alignment/mapping), динамическим обновлением моделей при дрейфе схем источников, а также с развитием механизмов объяснимости на основе provenance и графовых связей.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Cook M. Semantic Integration // Diffbot Knowledge Graph Glossary. – 2025. – URL: <https://blog.diffbot.com/knowledge-graph-glossary/semantic-integration/>.
2. Zuiev P., Kuchansky A., Biloshchytskyi A. et al. Development of Complex Methodology of Processing Heterogeneous Data in Intelligent Decision Support Systems // Eastern-European Journal of Enterprise Technologies. – 2020. – Vol. 4. – P. 14-23. – DOI: 10.15587/1729-4061.2020.209154.
3. Putrama I.M., Martinek P. Heterogeneous data integration: Challenges and opportunities // Data in Brief. – 2024. – Vol. 56. – Art. 110853. – DOI: 10.1016/j.dib.2024.110853.
4. Balaha F., Albinali H., Alrabiah H., Ali M., Bahroun Z. An analytical review of data integration for decision support in smart manufacturing // Decision Analytics Journal. – 2025. – Vol. 17. – Art. 100647. – DOI: 10.1016/j.dajour.2025.100647.
5. Rajabi E., Kafaie S. Knowledge Graphs and Explainable AI in Healthcare // Information. – 2022. – Vol. 13, No. 10. – Art. 459. – DOI: 10.3390/info13100459.
6. Rathore M.M., Ahmad A., Paul A., Rho S. Towards smart transportation: data integration and mining for large scale data // 2016 IEEE Region 10 Conference (TENCON Spring). – 2016. – DOI: 10.1109/TENCONSpring.2016.7519392.
7. R2RML: RDB to RDF Mapping Language. – W3C Recommendation, 27.09.2012. – URL: <https://www.w3.org/TR/r2rml/>.
8. JSON-LD 1.1. – W3C Recommendation, 16.07.2020. – URL: <https://www.w3.org/TR/json-ld11/>.
9. Gao F., Ali M. I., Mileo A. Semantic Discovery and Integration of Urban Data Streams // CEUR Workshop Proceedings. – 2014. – Vol. 1280. – P. 15-30. – URL: <https://ceur-ws.org/Vol-1280/>.
10. Hashem I.A.T., Chang V., Anuar N. B. et al. The role of big data in smart city // International Journal of Information Management. – 2016. – Vol. 36, No. 5. – P. 748-758. – DOI: 10.1016/j.ijinfomgt.2016.05.002.
11. Bamgboye O., Liu X., Cruickshank P. Semantic-Driven Approach for Validation of IoT Streaming Data in Trustable Smart City Decision-Making and Monitoring Systems // Big Data and Cognitive Computing. – 2025. – Vol. 9, No. 4. – Art. 108. – DOI: 10.3390/bdcc9040108.
12. Ribeiro M.B., Braghetto K.R.A Data Integration Architecture for Smart Cities // Anais do XXXVI Simpósio Brasileiro de Bancos de Dados (SBBd). – 2021. – DOI: 10.5753/sbbd.2021.18030.
13. OWL 2 Web Ontology Language Document Overview (Second Edition). – W3C Recommendation, 11.12.2012. – URL: <https://www.w3.org/TR/owl2-overview/>.
14. SPARQL 1.1 Query Language. – W3C Recommendation, 21.03.2013. – URL: <https://www.w3.org/TR/sparql11-query/> (дата обращения: 07.12.2025).
15. SPARQL 1.1 Query Language. – W3C Recommendation, 21.03.2013. – URL: <https://www.w3.org/TR/sparql11-query/> (дата обращения: 07.12.2025).
16. Shapes Constraint Language (SHACL). – W3C Recommendation, 20.07.2017. – URL: <https://www.w3.org/TR/shacl/>.
17. Михайлова С.С., Будаев Е.С. и др. Интеграция разнородных данных пространственного развития региона в СППР // Нелинейный мир. – 2025. – Т. 23, № 4. – С. 107-120.
18. Psyllidis A., Bozzon A., Bocconi S., Titos Bolivar C. A Platform for Urban Analytics and Semantic Data Integration in City Planning // Computer-Aided Architectural Design Futures. The Next City – New Technologies and the Future of the Built Environment / eds. G. Celani, D.M. Sperling, J.M.S. Franco. – Cham: Springer, 2015. – (Communications in Computer and Information Science; Vol. 527). – P. 21-36. – DOI: 10.1007/978-3-662-47386-3_2.
19. Song M. Research on intelligent decision support platform for tourism enterprises based on multi-source heterogeneous data fusion // Scientific Reports. – 2025. – Vol. 15, No. 1. – Art. 39810. – DOI: 10.1038/s41598-025-23486-x.
20. Осипов В.П., Сивакова Т.В., Судаков В.А., Нечаев Ю.И. Интеллектуальное ядро системы поддержки принятия решений // Препринты ИПМ им. М.В. Келдыша. – 2018. – № 205. – 23 с. – URL: <https://library.keldysh.ru/preprint.asp?id=2018-205>.

REFERENCES

1. Cook M. Semantic Integration, *Diffbot Knowledge Graph Glossary*, 2025. Available at: <https://blog.diffbot.com/knowledge-graph-glossary/semantic-integration/>.
2. Zuiev P., Kuchansky A., Biloshchytskyi A. et al. Development of Complex Methodology of Processing Heterogeneous Data in Intelligent Decision Support Systems, *Eastern-European Journal of Enterprise Technologies*, 2020, Vol. 4, pp. 14-23. DOI: 10.15587/1729-4061.2020.209154.
3. Putrama I.M., Martinek P. Heterogeneous data integration: Challenges and opportunities, *Data in Brief*, 2024, Vol. 56, Art. 110853. DOI: 10.1016/j.dib.2024.110853.

4. Balaha F., Albinali H., Alrabiah H., Ali M., Bahroun Z. An analytical review of data integration for decision support in smart manufacturing, *Decision Analytics Journal*, 2025, Vol. 17, Art. 100647. DOI: 10.1016/j.dajour.2025.100647.
5. Rajabi E., Kafaie S. Knowledge Graphs and Explainable AI in Healthcare, *Information*, 2022, Vol. 13, No. 10, Art. 459. DOI: 10.3390/info13100459.
6. Rathore M.M., Ahmad A., Paul A., Rho S. Towards smart transportation: data integration and mining for large scale data, 2016 *IEEE Region 10 Conference (TENCON Spring)*, 2016. DOI: 10.1109/TENCONSpring.2016.7519392.
7. R2RML: RDB to RDF Mapping Language. W3C Recommendation, 27.09.2012. Available at: <https://www.w3.org/TR/r2rml/>.
8. JSON-LD 1.1. W3C Recommendation, 16.07.2020. Available at: <https://www.w3.org/TR/json-ld11/>.
9. Gao F., Ali M. I., Mileo A. Semantic Discovery and Integration of Urban Data Streams, *CEUR Workshop Proceedings*, 2014, Vol. 1280, pp. 15-30. Available at: <https://ceur-ws.org/Vol-1280/>.
10. Hashem I.A.T., Chang V., Anuar N. B. et al. The role of big data in smart city, *International Journal of Information Management*, 2016, Vol. 36, No. 5, pp. 748-758. DOI: 10.1016/j.ijinfomgt.2016.05.002.
11. Bamgboye O., Liu X., Cruickshank P. Semantic-Driven Approach for Validation of IoT Streaming Data in Trustable Smart City Decision-Making and Monitoring Systems, *Big Data and Cognitive Computing*, 2025, Vol. 9, No. 4, Art. 108. DOI: 10.3390/bdcc9040108.
12. Ribeiro M.B., Braghetto K.R.A Data Integration Architecture for Smart Cities, *Anais do XXXVI Simpósio Brasileiro de Bancos de Dados (SBBd)*, 2021. DOI: 10.5753/sbbd.2021.18030.
13. OWL 2 Web Ontology Language Document Overview (Second Edition). W3C Recommendation, 11.12.2012. Available at: <https://www.w3.org/TR/owl2-overview/>.
14. SPARQL 1.1 Query Language. W3C Recommendation, 21.03.2013. Available at: <https://www.w3.org/TR/sparql11-query/> (accessed 07 December 2025).
15. SPARQL 1.1 Query Language. W3C Recommendation, 21.03.2013. Available at: <https://www.w3.org/TR/sparql11-query/> (accessed 07 December 2025).
16. Shapes Constraint Language (SHACL). W3C Recommendation, 20.07.2017. Available at: <https://www.w3.org/TR/shacl/>.
17. Mikhaylova S.S., Budaev E.S. i dr. Integratsiya raznorodnykh dannykh prostranstvennogo razvitiya regiona v SPPR [Integration of heterogeneous data on the spatial development of a region in decision-making systems], *Nelineyny mir [Nonlinear World]*, 2025, Vol. 23, No. 4, pp. 107-120.
18. Psyllidis A., Bozzon A., Bocconi S., Titos Bolivar C. A Platform for Urban Analytics and Semantic Data Integration in City Planning, *Computer-Aided Architectural Design Futures. The Next City – New Technologies and the Future of the Built Environment*, eds. G. Celani, D.M. Sperling, J.M.S. Franco. Cham: Springer, 2015. – (Communications in Computer and Information Science; Vol. 527), pp. 21-36. DOI: 10.1007/978-3-662-47386-3_2.
19. Song M. Research on intelligent decision support platform for tourism enterprises based on multi-source heterogeneous data fusion, *Scientific Reports*, 2025, Vol. 15, No. 1, Art. 39810. DOI: 10.1038/s41598-025-23486-x.
20. Osipov V.P., Sivakova T.V., Sudakov V.A., Nechaev Yu.I. Intellektual'noe yadro sistemy podderzhki prinyatiya resheniy [Intelligent core of the decision support system], *Preprinty IPM im. M.V. Keldysha* [Preprints of the Keldysh Institute of Problems of Applied Mathematics], 2018, No. 205, 23 p. Available at: <https://library.keldysh.ru/preprint.asp?id=2018-205>.

Гапочка Владислав Валерьевич – Кубанский государственный университет; e-mail: gapochka2000@mail.ru; тел.: +89883667270; кафедра вычислительных технологий; аспирант.

Полупанова Елена Евгеньевна – Кубанский государственный университет; e-mail: jienka@mail.ru; г. Краснодар, Россия; тел.: +79284013301.; кафедра вычислительных технологий; к.т.н.; доцент; ORCID: <https://orcid.org/0000-0002-0364-1132>.

Gapochka Vladislav Valerievich – Kuban State University; e-mail: gapochka2000@mail.ru; Krasnodar, Russia; phone: +89883667270, the Department of Computational Technologies; graduate student.

Polupanova Elena Evgenievna – Kuban State University; e-mail jienka@mail.ru; Krasnodar, Russia; phone: +79284013301; the Department of Computational Technologies; cand. of eng. sc.; associate professor; ORCID: <https://orcid.org/0000-0002-0364-1132>.

Н.О. Лужевский, В.Ф. Лубенцов, Е.В. Лубенцова**СИНТЕЗ РОБАСТНОЙ САУ С ДИНАМИЧЕСКИМ КОМПЕНСАТОРОМ
И НАБЛЮДАТЕЛЕМ СОСТОЯНИЯ С КОРРЕКТИРУЮЩИМИ ОБРАТНЫМИ
СВЯЗЯМИ НА ОСНОВЕ СИГМОИДНЫХ ФУНКЦИЙ**

Изложена методика анализа робастной САУ для линейных динамических объектов управления с запаздыванием при неопределенности информации о параметрах математической модели и помехах в измеряемых сигналах. В системе реализованы широко распространенный ПИД-регулятор, динамический компенсатор (ДК) инерционной части объекта управления, реализуемый с использованием оценок переменных состояния, получаемых на основе наблюдателя состояния (НС). Отмечено, что для асимптотической устойчивости регулятора и наблюдателя можно воспользоваться любым критерием устойчивости, но для обеспечения максимальной степени устойчивости и требуемых показателей качества переходного процесса удобно воспользоваться критерием максимальной степени устойчивости. В данной работе предлагается вместо производных, получаемых дифференцированием, использовать в ДК инерционности объекта оценки переменных состояния объекта, получаемых с помощью НС. Другим отличием от известных является использование в корректирующих обратных связях НС сигмоидных функций и реализация дополнительного к основному воздействию на основе ошибки оценивания, компенсирующего внешнее возмущение на входе объекта. При параметрическом синтезе НС исследовано влияние возмущений и помех на качество функционирования САУ промышленным интервально-заданным объектом при различных его параметрах. Рассчитан робастный типовой ПИД-регулятор с параметрами, оптимальными по критерию максимальной степени устойчивости, с учетом компенсации инерционной части объекта. Для этого был реализован ДК с помощью параметров номинальной (расчетной) модели объекта с наилучшим сочетанием параметров, полученной на основе интервальной модели объекта управления с запаздыванием. Проведенными исследованиями установлено, что структура САУ с типовым ПИД-регулятором с последовательным ДК инерционности объекта и НС с корректирующими обратными связями на основе сигма-функций обеспечивает простоту методики синтеза САУ, робастной к изменениям параметров объекта и действию неизмеряемых возмущений и неконтролируемых помех.

Робастное управление биореактором; типовой ПИД-регулятор; динамический компенсатор; наблюдатель состояния; сигма-функция.

N.O. Luzhevsky, V.F. Lubentsov, E.V. Lubentsova**SYNTHESIS OF A ROBUST ACS WITH A DYNAMIC COMPENSATOR
AND A STATE OBSERVER WITH CORRECTIVE FEEDBACK BASED ON SIGMOID
FUNCTIONS**

A methodology for analyzing a robust automatic control system for linear dynamic control objects with a delay in the uncertainty of information about the parameters of the mathematical model and interference in the measured signals is described. The system implements a widely used PID controller and a dynamic compensator (DC) for the inertial part of the control object, implemented using estimates of state variables obtained on the basis of a state observer (SO). It is noted that any stability criterion can be used for the asymptotic stability of the controller and observer, but to ensure the maximum degree of stability and the required quality indicators of the transient process, it is convenient to use the maximum degree of stability criterion. In this paper, instead of derivatives obtained by differentiation, it is proposed to use in the dynamic compensator of the inertia of the object estimates of the state variables of the object obtained with the help of a state observer. Another difference from the known ones is the use of sigmoid functions in the corrective feedback of the state observer and the implementation of an additional effect to the main one based on the estimation error, compensating for external disturbance at the input of the object. The parametric synthesis of a state observer was used to study the impact of disturbances and noise on the performance of an automated control system (ACS) for an industrial interval-defined plant with various parameters. A robust typical PID controller with optimal parameters for maximum stability was calculated, taking into account the compensation of the inertial part of the object. For this purpose, a dynamic compensator (DC) was implemented using the parameters of a nominal (calculated) model of the object with the worst combination of parameters obtained on the basis of the interval model of the control object with a delay. The conducted research has established that the structure of the ACS with a typical PID

controller with a sequential dynamic compensator for the inertia of the object and a state observer with corrective feedback based on sigma functions ensures the simplicity of the ACS synthesis methodology that is robust to changes in the parameters of the object and the action of unmeasured disturbances and uncontrolled interference.

Robust bioreactor control; typical PID controller; dynamic compensator; state observer; sigma function.

Введение. При синтезе САУ системными объектами, к которым можно отнести сложные технические и биотехнологические системы, используются различные методы современной теории управления. К ним следует отнести методы робастного управления, динамической компенсации, гарантирующего управления, интервальных оценок и их комбинации [1]. Ряд этих методов для структурно-параметрического синтеза робастных систем управления, в частности для реализации систем с динамическими компенсаторами (ДК) инерционностей объекта управления (ОУ), требует точного знания всех переменных состояний. Однако на практике измерить все переменные состояния не представляется возможным. Одна из первых работ синтеза систем управления с ДК для обеспечения робастности к инерционности объекта связана с использованием метода обратных операторов [2], либо дифференциальных уравнений, описывающих объект. Применение производных в ДК может приводить к усилению помех в измеряемых сигналах. В этом случае получить информацию о производных в модели компенсатора, не используя многократное дифференцирование измеряемой выходной переменной, в современных системах возможно на основе оценок переменных состояний с помощью наблюдателей состояния (НС). В случае приближенных моделей, получаемых зачастую по экспериментальным данным и в условиях ряда допущений о квазистационарности, линейности или линеаризации и др., требуется обеспечение робастности не только к параметрической неопределенности модели объекта и действующим на систему внешним возмущениям, но и к шумам в канале измерения. С учетом этого выбор значений коэффициентов обратной связи НС должен быть осуществлен с учетом реального влияния помех на качество САУ в условиях, когда отсутствуют возможности точного измерения производных и получения информации о возмущениях и шумах в измеряемом сигнале на выходе объекта, а используются только оценки переменных состояния управляемого объекта.

Как известно [3], при больших изменениях параметров объекта методы теории чувствительности для синтеза робастной системы управления непригодны. В этом случае одним из подходов является синтез робастной системы с постоянными параметрами регуляторов и компенсирующих устройств, полученных с помощью номинальных (расчетных) моделей объекта с различными параметрами, например, с наихудшим их сочетанием на основе интервальной модели объекта [4]. Этот подход является более общим по сравнению с методами, основанными на оценке чувствительности. В связи с тем, что при этом подразумеваются большие изменения параметров и возмущений, включая помехи в измеряемых сигналах, то регулятор (система) должен быть робастным (нечувствительным) не только к большим изменениям параметров объекта, но и к действию неконтролируемых помех (шумов) в измеряемых сигналах. Поэтому рассматриваемая в данной работе задача обеспечения робастности САУ с ДК и НС к изменениям параметров объекта и к действию неконтролируемых помех в сигналах измерения, синтезированной с помощью интервальной модели объекта, является актуальной.

Постановка задачи. В работе [5] отмечено, что «для обеспечения высокого качества оценивания требуется организовать скользящий режим, близкий к идеальному, при котором изображающая точка системы, записанной относительно ошибок наблюдения, движется по многообразию скольжения, совершая колебания с бесконечно большой частотой и бесконечно малой амплитудой. При большой, но конечной частоте переключения возникает реальный скользящий режим ...», в котором систему можно считать эквивалентной адаптивной, т.е. приобретающей адаптивные свойства, близкие к свойствам систем с сигнальной адаптацией. Однако учитывая тот факт, что регулируемые техноло-

гические объекты являются инерционными, а каналы регулирования содержат большое запаздывание и исполнительные устройства с ограниченной скоростью перемещения регулирующего органа, скользящий режим при ограниченной частоте переключений может быть трудноосуществим либо нереализуем. Несмотря на это, следует отметить, что поскольку оценка переменных состояний в наблюдателе осуществляется путем интегрирования уравнений наблюдателя, то искомые переменные отфильтровываются от высокочастотных помех, входящих в измеряемый выходной сигнал $y = x + \xi$ (x – переменная состояния, ξ – помеха) [6]. Следует также ожидать, что использование сигмоидных функций в корректирующих обратных связях наблюдателя способствует повышению темпа оценивания и уменьшению ошибок, связанных с неточностью математической модели объекта. Обычно быстродействие НС делают на 20-50 % больше, чем быстродействие контура управления с регулятором объекта [6].

В связи с этим становится актуальной альтернатива решения задачи синтеза САУ, робастной к изменениям параметров объекта и действию неконтролируемых помех, в структуре которой используется сочетание типового ПИД-регулятора, последовательного ДК инерционности объекта, реализуемого на основе переменных состояния НС с корректирующими обратными связями на основе сигмоидных функций с насыщением вне скользящего режима.

Материалы и методы исследования. Одним из эффективных средств повышения точности работы САУ в установившемся режиме является компенсация инерционности ОУ и действующих на него внешних возмущений. Условия реализации таких видов компенсаций во многом аналогичны. Для полной компенсации инерционности ОУ и возмущения в САУ необходимо вводить сигналы, пропорциональные первой и высшим производным от выходной координаты. Полная компенсация возмущающего воздействия считается достигнута, когда после завершения переходного процесса регулируемая (или управляемая) величина и ошибка системы не зависят от этого возмущения. Полная компенсация инерционности ОУ считается достигнута, когда все полученные реализуемые производные не приводят к дополнительным движениям составляющих переходного процесса в САУ и не снижают требуемую точность. В реальных системах можно точно получить лишь первую производную, все последующие производные могут быть получены лишь приближенно. Это приводит к достижению не полной (абсолютной), а частичной компенсации.

Однако и частичная инвариантность может быть весьма полезной в смысле повышения точности САУ, если имеется возможность измерения возмущений. Но во многих случаях непосредственное измерение возмущений или затруднительно, или практически невозможно из-за отсутствия датчиков.

Недостаточность измерительной информации обуславливает состояние неопределенности, что является характерной особенностью многих процессов и производств микробиологической, химико-фармацевтической, пищевой и других отраслей промышленности. Ключевым аспектом разрешения этой ситуации является использование переменных состояний. Это направление интенсивно развивается, о чем свидетельствует нарастающий поток публикаций. Создаются и совершенствуются различные теории, методы и подходы. В существующих разработках следует отметить применение метода интервальных оценок, метода синтеза робастных регуляторов, а также методов, основанных на использовании НС и др. Обзор некоторых из них приведен в работе [1]. Каждый из методов характеризуется своими возможностями, достоинствами и недостатками, областями применения. Причем каждый метод длительное время использовался как самостоятельный, вне комбинации с другими. В этом аспекте следует отметить, несмотря на появление новых алгоритмов управления, широкое использование при управлении технологическими и техническими объектами типовых законов регулирования (П, ПИ, ПИД, ПД). В [7] отмечено, что около 95 % используемых на практике регуляторов составляют регуляторы с ПИД-законом управления. Существует большое количество методов настройки типовых регуляторов [2–4], использующих эти законы регулирования, которые позволяют полу-

чить заданные качественные показатели системы. В статье [8] отмечено, что непрекращающийся поток публикаций по настройке типовых регуляторов свидетельствует об актуальности этой проблемы. Объясняется это тем, что при использовании систем с применением типовых законов регулирования достигается хорошее качество регулирования и имеется возможность совершенствования систем с применением ПИД-регуляторов различных модификаций при возрастании требований со стороны промышленных объектов [9–12]. Однако использование отдельных методов и традиционных законов регулирования оказывается недостаточно эффективным, особенно при синтезе САУ сложными объектами и возрастающих требованиях к качеству их функционирования в условиях параметрической неопределенности и внешних возмущениях. Перспективным является использование ПИД-закона в комбинации с указанными методами, что является предметом рассмотрения в данной работе. Как известно из теории автоматического управления, ключевым моментом при использовании типовых законов регулирования является наличие номинальной (расчетной) модели объекта [13]. Возможным подходом к решению задачи выбора номинальной (расчетной) модели в САУ с типовыми регуляторами может быть использование в расчетной (номинальной) модели объекта граничных значений параметров, соответствующих их наиболее неблагоприятному сочетанию [14], с последующим синтезом настроек регулятора по критерию максимальной степени устойчивости, обеспечивающим распределение полюсов, наиболее близко расположенных к мнимой оси [9, 15]. Выбор подходящего распределения полюсов замкнутой САУ является важным этапом при обеспечении заданных показателей качества переходного процесса: длительности переходного процесса, перерегулирования и др.

Во многих случаях необходимое повышение робастных свойств САУ может быть достигнуто определенным изменением распределения полюсов на основе использования в расчетной модели неблагоприятного сочетания параметров интервальной модели. Оценку возможности и целесообразности такой методики наиболее удобно проводить на основе анализа переходных процессов САУ, полученных при параметрах регулятора для максимальной степени устойчивости, которые дают наиболее полную картину влияния неблагоприятного сочетания параметров интервальной модели на устойчивость системы.

Идея построения управления с использованием граничных значений параметров номинальной (расчетной) модели объекта состоит в следующем. При синтезе параметров регулятора актуальной является проблема неопределенности в модели (неизвестные параметры и/или внешние возмущения). В этом случае использование точных значений параметров модели затруднительно по следующим причинам:

- ♦ синтез традиционными методами является слишком громоздким, поскольку выполнить полный перебор вариантов модели при точных значениях параметров из интервала их изменения затруднено из-за отсутствия универсальности этих методов для случая неопределенности;
- ♦ система может оказаться недостаточно робастной, поскольку точечные модели (точечные модели предоставляют информацию только о параметрах объекта в отдельные моменты времени, описывая его набором моделей), особенно для нестационарных объектов, оказываются приближенными;
- ♦ синтезировать один регулятор, робастный во всем интервале параметров объекта, используя перебор n точечных моделей и не прибегая к адаптивным методам, невозможно из-за отсутствия рекомендаций по выбору оптимального варианта.

В данной работе основное внимание уделяется разработке методики, позволяющей при соблюдении прочих условий учитывать требования робастности, т. е. нечувствительности системы к изменению параметров, при которых выполняются заданные показатели качества переходных процессов во всех интервалах исследования. В качестве номинальной (расчетной) модели объекта используется передаточная функция в виде линейного инерционного звена второго порядка с запаздыванием. При этом предполагается, что параметры передаточной функции и запаздывание могут варьироваться в априори

известных интервалах, а их граничные значения известны. В основе методики лежит использование расчетной модели объекта с граничными значениями параметров, соответствующими наиболее неблагоприятному их сочетанию. Этот метод ориентирован на нахождение корней, наиболее близко расположенных к границе устойчивости, обеспечивая простоту процедуры синтеза и применимость для широкого класса объектов.

Другим ключевым моментом в задаче повышения качества управления является применение ПИД-регулятора с последовательным динамическим компенсатором инерционности объекта. В работе [16] отмечается, что синтез компенсирующего устройства с передаточной функцией, обратной передаточной функции объекту управления, представляет неразрешимую с абсолютной точностью задачу, поскольку в такой структуре должны быть выполнены операции точного дифференцирования [16]. Поэтому в современной теории управления существует ряд методов приближенного решения подобного рода некорректных задач. Большинство из них строится по методу малого параметра, позволяющего получить с некоторой ошибкой физически реализуемую структуру компенсирующего устройства $W_k(p)$, обратную по отношению к передаточной функции объекта $W_{об}(p)$. Основной недостаток такого подхода состоит в низкой точности реализации обратной передаточной функции при заданном запасе устойчивости.

В [17] предлагается замена идеальных дифференцирующих звеньев реальными. Например, применение не идеального, а физически реализуемого дифференцирующего звена с малой постоянной времени (например, равной 0,1) незначительно изменяет процессы регулирования. Следовательно, при реализации ДК на основе обратной модели только инерционной части передаточной функции объекта (т.е. без учета запаздывания) имеется возможность компенсации, реализуемой практически. Вместе с тем в [19, 20] отмечается, что многие методы синтеза регуляторов в неявной форме реализуют принцип динамической компенсации. В качестве одного из методов приведем решение задачи управления детерминированным объектом с запаздыванием на основе метода динамической компенсации при наличии неконтролируемого возмущения [21]. Многие промышленные объекты управления могут быть описаны моделью вида

$$W_o(p) = k \cdot \exp(-\tau p) / (Tp + 1), \quad (1)$$

где k – коэффициент передачи; T – постоянная времени; τ – запаздывание.

Для достижения максимального быстродействия необходимо подать на объект максимально возможное управляющее воздействие с целью получения максимальной скорости переходного процесса, выдержать время T_u , а затем уменьшить его до величины возмущения F , но с обратным знаком. Для решения задачи требуется определить длительность T_u управляющего воздействия и величину возмущения F .

Переходный процесс на выходе объекта (1) при действии возмущения F ступенчатого вида описывается уравнением

$$x_{\text{вых}}(t) = k \cdot F \cdot [1 - \exp(-t/T)]. \quad (2)$$

Переходный процесс при подаче управляющего воздействия максимальной величины U_{max} на интервале времени $T_u = t - \tau$ описывается уравнением

$$x_{\text{вых}}^*(t) = k \cdot U_{\text{max}} \cdot [1 - \exp(-T_u/T)]. \quad (3)$$

Переключение управляющего воздействия с величины U_{max} на величину $U = -F$ необходимо производить в момент равенства модулей ординат кривых $x_{\text{вых}}(t)$ и $x_{\text{вых}}^*(t)$. Выбрав $U_{\text{max}} = kF$, получим

$$U_{\text{max}} \cdot [1 - \exp(-(T_u + \tau)/T)] = k \cdot U_{\text{max}} \cdot [1 - \exp(-T_u/T)]. \quad (4)$$

Отсюда находим

$$T_u = T \cdot \ln[(k - \exp(-\tau/T))/(k - 1)]. \quad (5)$$

Для определения величины возмущения F на входе объекта по изменению выходной величины объекта $X_{\text{вых}}(t)$ используется обратная модель передаточной функции объекта (1) без запаздывания, т.е. используется оценка F на основе оператора $\hat{F}(p) = W_M^{-1}(p) \cdot X_{\text{вых}}(p)$ с последующей компенсацией возмущения F .

Таким образом, если на объект управления действует неконтролируемое возмущение F , аддитивно влияющее на выходную величину, то информацию о приведённом ко входу возмущении при известной динамике канала возмущения можно получить, используя метод динамической компенсации на основе обратной модели передаточной функции объекта управления. На основе этой информации управляющее воздействие после определения величины возмущающего воздействия $\hat{F}$ реализуется с помощью компенсации F в виде

$$U(t) = U_X(t)|_{t=T_u} + U_F(t)|_{t>T_u} = U_{\max} - \hat{F}, \quad (6)$$

где $U_{\max}$ – управляющее воздействие, определяющее переходный процесс системы при $\hat{F} = 0$; U_F – управляющее воздействие, компенсирующее влияние $\hat{F}$ на систему.

В системах, включающих ДК инерционности объекта и дополнительные составляющие, компенсирующие действие возмущений, робастность переменных состояния по отношению к неопределенностям может быть обеспечена с помощью нелинейных корректирующих обратных связей в НС и организации дополнительного канала компенсации возмущения на основе их оценок. Практически значимый метод оценивания переменных состояния и возмущения, не требующий получения их производных и динамической модели, состоит в применении НС с корректирующими обратными связями на основе сигмоидных функций [5, 22].

В данной работе предлагается вместо производных, получаемых дифференцированием, использовать в ДК инерционности объекта оценки переменных состояния объекта, получаемых с помощью НС. Другим отличием от известных является использование в корректирующих обратных связях НС сигмоидных функций и реализация дополнительного к основному воздействию на основе ошибки оценивания, компенсирующего внешнее возмущение на входе объекта.

В общем случае сигмоидная функция определяется выражением:

$$\sigma(\varepsilon, \lambda) = \frac{1}{1 + \exp(-\lambda(\varepsilon - a))} - \frac{1}{1 + \exp(\lambda(\varepsilon + a))}, \quad (7)$$

где ε – сигнал обратной связи в НС; $\lambda > 0$ – коэффициент, характеризующий крутизну логистической функции, усиливающей слабые сигналы ε (параметр наклона сигмоидальной функции); $l = \text{const}$ – максимальное значение пороговой функции; a – значение смещения (в отдельных случаях может отсутствовать).

Если значение коэффициента λ велико, то график сигмоидной функции в общем виде приближается к пороговой функции (рис. 1).

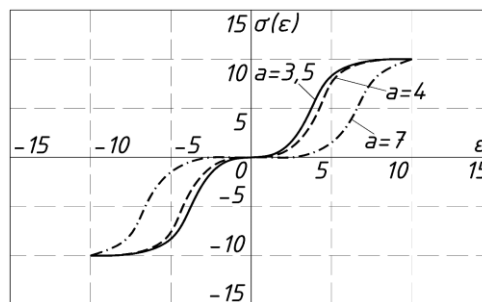


Рис. 1. График сигмоидной функции при $l = 10$ со смещением $a = 3,5; 4; 7$

Проведем численное моделирование САУ с НС, коэффициенты обратной связи в котором являются непрерывными функциями сигмоидного типа без смещения и изменяются в соответствии с выражениями:

$$\sigma_1(e) = \frac{l_1}{1 + \exp(-\lambda e)} - \frac{l_1}{1 + \exp(\lambda e)}; \quad (8)$$

$$\sigma_2(e) = \frac{l_2}{1 + \exp(-\lambda e)} - \frac{l_2}{1 + \exp(\lambda e)}; \quad (9)$$

$$\sigma_3(e) = \frac{l_3}{1 + \exp(-\lambda e)} - \frac{l_3}{1 + \exp(\lambda e)}, \quad (10)$$

где e – ошибка оценивания; λ – коэффициент, характеризующий крутизну логистической функции; $\sigma_1(e)$, $\sigma_2(e)$, $\sigma_3(e)$ – коэффициенты обратной связи НС сигмоидного типа; l_1 , l_2 , l_3 – постоянные коэффициенты.

Структурная схема системы управления с ПИД-регулятором, ДК на основе обратной модели инерционности объекта управления $W_m^{-1}(p)$ и НС на основе номинальной модели объекта управления $W_m(p) \cdot e^{-p\tau}$ представлена на рис. 2.

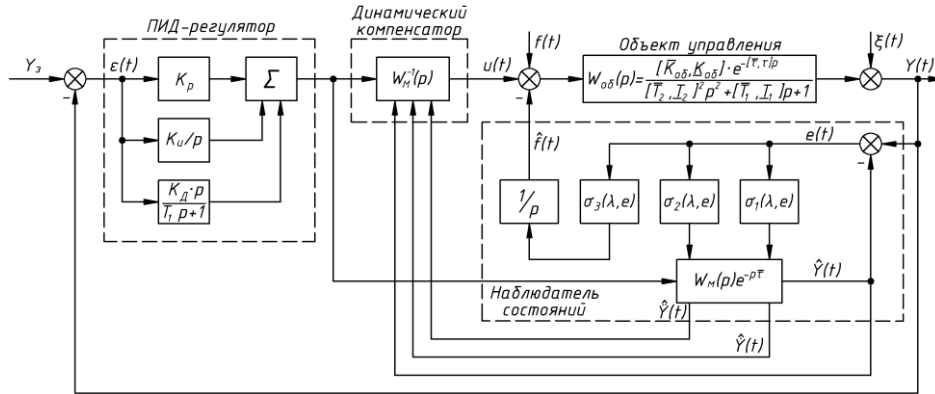


Рис. 2. Структурная схема системы управления с ПИД-регулятором, ДК на основе обратной модели инерционности объекта управления и НС: $W_m(p) \cdot e^{-p\tau}$ – номинальная модель объекта управления; $W_m^{-1}(p)$ – передаточная функция ДК; p – оператор дифференцирования; $Y_3(t)$ – заданное значение регулируемой переменной $Y(t)$; $\hat{Y}(t)$ – оценка регулируемой переменной $Y(t)$; $\hat{\dot{Y}}(t)$ – оценка первой производной $Y(t)$; $\hat{\ddot{Y}}(t)$ – оценка второй производной $Y(t)$; $e(t)$ – ошибка оценивания; $u(t)$ – управляющее воздействие; $f(t)$, $\hat{f}(t)$ – возмущение и его оценка; $\xi(t)$ – помеха

НС с обратными связями на основе выражений (8), (9), (10) реализует алгоритм непрерывного во времени оценивания производных. Для решения задачи оценивания с помощью НС с учетом рекомендаций [5, 22] приняты следующие параметры корректирующих сигмоидальных воздействий в обратных связях НС: $l_1 = 0,1778$; $l_2 = 0,00034$; $l_3 = 0,0638$; $\lambda = 1,0$. Согласно методу синтеза наблюдателей состояний [5] в наблюдателе не используется динамическая модель возмущения. Данный наблюдатель является робастным, так как величины корректирующих воздействий в обратных связях наблюдателя реализуются на основе интервальной модели объекта управления и не требуют точной математической модели объекта. Это позволяет не увеличивать порядок наблюдателя для получения оценки возмущающего воздействия, не требует реализации большой частоты переключений и дает возможность исключить большие динамические отклонения зашумленной регулируемой переменной при одновременном действии задающего и возмущающего сигналов в замкнутой системе с наблюдателем. Этому способствует реализация кор-

ректирующих воздействий в обратной связи наблюдателя, которые являются переменными в ходе работы наблюдателя и реализуются на основе S-образных непрерывных функций с насыщением. Значения коэффициента λ сигмоидных функций определены из обеспечения условий приближения характеристики сигмоидной функции к квазирелейной для усиления слабых сигналов либо к линейной в зависимости от ошибки оценивания $e(t)$.

В качестве иллюстрации разработанной методики рассмотрим задачу управления биообъектом по каналу регулирования растворенного в биореакторе кислорода (pO_2), описываемого интервальной моделью 2-го порядка с запаздыванием в следующем виде

$$W(p) = \frac{[\bar{K}_{об}, \underline{K}_{об}] e^{-[\bar{\tau}, \underline{\tau}]p}}{[\bar{T}_2, \underline{T}_2]^2 p^2 + [\bar{T}_1, \underline{T}_1] p + 1},$$

где $\underline{K}_{об}, \bar{K}_{об}, \underline{T}_1, \bar{T}_1, \underline{T}_2, \bar{T}_2, \underline{\tau}, \bar{\tau}$ – нижняя и верхняя граница коэффициента передачи объекта $K_{об}$, постоянных времени T_1, T_2 и запаздывания объекта τ соответственно.

Отметим, что впервые интервальные наблюдатели для неопределенных биотехнологических систем, предназначенные для мониторинга биореактора, были предложены в работе [23]. Непрерывные наблюдатели для неопределенных биотехнологических моделей впервые были рассмотрены в работе [24]. Разработка методов оценки состояния для неопределенных линейных систем изложена в работе [25]. В работе [26] рассмотрен метод построения интервального наблюдателя, позволяющего получить область оценок переменных состояния процесса биосинтеза – концентрации клеток и концентрации субстрата – и их интервальные оценки, ограниченные верхней и нижней границами. Такой характер изменения важных параметров состояния процесса подтверждает наличие различного рода неопределенностей в динамике биосистемы и говорит о необходимости решения задачи управления процессом насыщения среды кислородом с помощью робастного автоматического регулирования pO_2 .

В данной работе решена задача синтеза САУ процессом насыщения среды культивирования кислородом. Исследование проведено для четырех вариантов модели объекта $W_{M1}(p), W_{M2}(p), W_{M3}(p), W_{M4}(p)$, представленных в табл. 1, при следующих граничных значениях параметров интервальной модели объекта управления: $K_{об\max} = 0,375$, $\tau_{\max} = 7,5$ мин, $T_{1\max} = 24,96$ мин, $T_{1\min} = 17,62$ мин, $T_{2\max} = 14,8$ мин, $T_{2\min} = 13,02$ мин.

Таблица 1

Значения настроек ПИД-регулятора максимальной степени устойчивости J , соответствующих неблагоприятным значениям параметров модели объекта $W_{M1}(p), W_{M2}(p), W_{M3}(p), W_{M4}(p)$ по каналу регулирования pO_2

Номер модели	Номер САУ	$K_{об\max}$	$\tau_{\max}$	T_1	T_2	J	K_{Π}	K_{Π}	K_{Δ}
$W_{M1}(p)$	1	0,375	7,5	24,96	14,8	0,094	3,339	0,125	29,665
$W_{M2}(p)$	2	0,375	7,5	17,62	14,8	0,086	1,934	0,089	28,463
$W_{M3}(p)$	3	0,375	7,5	24,96	13,02	0,102	3,413	0,136	23,011
$W_{M4}(p)$	4	0,375	7,5	17,62	13,02	0,094	1,884	0,094	21,151

Характеристическое уравнение НС должно быть асимптотически устойчивым [6]. Поэтому для асимптотической устойчивости можно воспользоваться любым критерием устойчивости. Для того, чтобы обеспечить заданный переходный процесс, целесообразно воспользоваться критерием максимальной степени устойчивости [9], с помощью которого рассчитываются коэффициенты закона управления. При этом под степенью устойчивости понимается расстояние от мнимой оси до ближайшего корня или до ближайшей пары комплексно-сопряженных корней.

При вариантах модели объекта с постоянными времени из табл. 1 рассчитаны по критерию максимальной степени устойчивости $J = -\min \max \operatorname{Re} p_k(r_i)$ (p_k – корни характеристического уравнения) параметры настройки ПИД-регулятора r_i ($i = 1, 2, 3$) [9]. Для максимальной степени устойчивости $J = 0,086$ определены $K_{\Pi} = 1,934$; $K_{\Pi} = 0,089$;

$K_d = 28,463$ (вторая строка в табл. 1). Переходные процессы в САУ (рис. 3, 4) получены при действии помехи в измеряемом сигнале на выходе объекта в диапазоне дисперсии помехи $D_{\text{вх}}(\xi) = [0,01, 10]$ и при ступенчатом изменении возмущения $f(t)$ на входе объекта в установившемся состоянии

$$f(t) = \begin{cases} 0,5 & \text{при } 0 \leq t \leq 200, \\ 1,5 & \text{при } 200 < t \leq 400, \\ 3,0 & \text{при } 400 < t \leq 700, \\ 1,0 & \text{при } 700 < t \leq 1100. \end{cases}$$

Результаты оценки флуктуации регулируемой переменной $Y(t)$ и ее оценки $\hat{Y}(t)$ при минимальной и максимальной значениях дисперсии помехи $D_{\text{вх}}(\xi)$ приведены в табл. 2 и 3, а переходные процессы представлены на рис. 3 и 4.

Таблица 2

Значения дисперсии и среднеквадратического отклонения (СКО) регулируемой переменной $Y(t)$ и ее оценки $\hat{Y}(t)$ от дисперсии помехи $D_{\text{вх}}(\xi) = 0,01$

Номер модели	Дисперсия помехи $D_{\text{вх}}(\xi)$	Дисперсия регулируемой переменной $Y(t)$	СКО регулируемой переменной $Y(t)$	Дисперсия оценки $\hat{Y}(t)$ регулируемой переменной	СКО оценки $\hat{Y}(t)$ регулируемой переменной
$W_{M1}(p)$	0,01	2,297	1,516	2,191	1,480
$W_{M2}(p)$	0,01	2,085	1,444	2,119	1,456
$W_{M3}(p)$	0,01	2,227	1,492	2,188	1,479
$W_{M4}(p)$	0,01	2,001	1,415	2,119	1,456

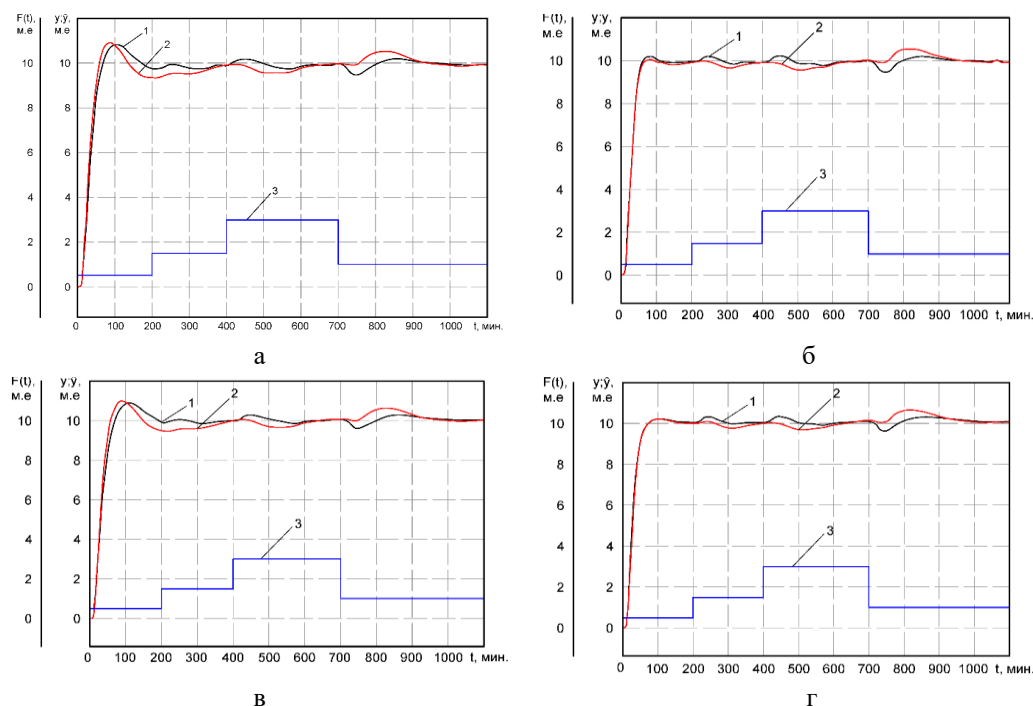


Рис. 3. Переходные процессы в САУ с ПИД-регулятором и моделью объекта $W_{M1}(p)$, $W_{M2}(p)$, $W_{M3}(p)$, $W_{M4}(p)$ (а, б, в, г соответственно) по каналу регулирования pO_2 $D_{\text{вх}}(\xi) = 0,01$

Таблица 3

Значения дисперсии и среднеквадратического отклонения (СКО) регулируемой переменной $Y(t)$ и ее оценки $\hat{Y}(t)$ от дисперсии помехи $D_{\text{вх}}(\xi) = 10$

Номер модели	Дисперсия помехи $D_{\text{вх}}(\xi)$	Дисперсия регулируемой переменной $Y(t)$	СКО регулируемой переменной $Y(t)$	Дисперсия оценки $\hat{Y}(t)$ регулируемой переменной	СКО оценки $\hat{Y}(t)$ регулируемой переменной
$W_{M1}(p)$	10	2,288	1,513	2,217	1,489
$W_{M2}(p)$	10	2,087	1,445	2,148	1,466
$W_{M3}(p)$	10	2,219	1,49	2,226	1,492
$W_{M4}(p)$	10	2,004	1,416	2,164	1,471

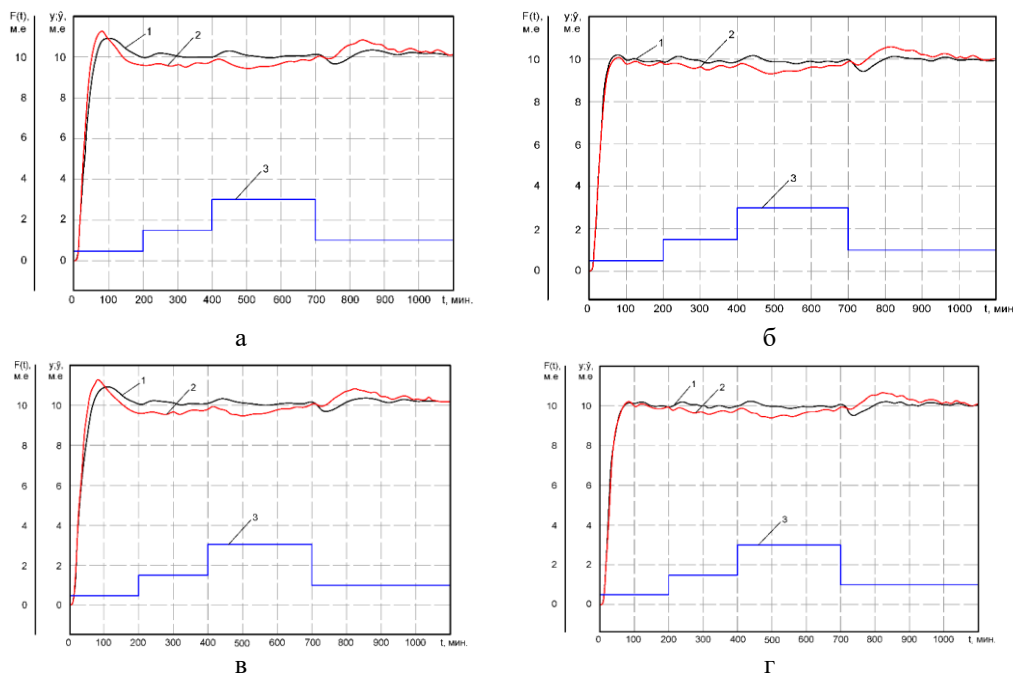


Рис. 4. Переходные процессы в САУ с ПИД-регулятором и моделью объекта $W_{M1}(p)$, $W_{M2}(p)$, $W_{M3}(p)$, $W_{M4}(p)$ (а,б,в,г соответственно) по каналу регулирования pO_2 $D_{\text{вх}}(\xi) = 10$

В качестве показателя помехоустойчивости использована величина среднеквадратического отклонения (СКО) координат состояния на выходе объекта управления и на выходе НС, наиболее подверженных влиянию помех в установившемся режиме. В каждом интервале был вычислен коэффициент ослабления $K_{\text{осл}} = D_{\text{вх}}(\xi)/D_{\text{вых}}(\xi)$ [27]. С помощью $K_{\text{осл}}$ можно дать оценку поведению объекта относительно входа/выхода при различных моделях объекта и различном уровне помех в измеряемом сигнале. С увеличением дисперсии помехи на входе $D_{\text{вх}}(\xi)$ коэффициент ослабления $K_{\text{осл}}$ для всех моделей объекта M_1 , M_2 , M_3 , M_4 с передаточными функциями $W_{M1}(p)$, $W_{M2}(p)$, $W_{M3}(p)$, $W_{M4}(p)$, изменяется равномерно, обеспечивая примерно одинаковое значение относительного коэффициента ослабления $K_{\text{осл отн}}^* = K_{\text{осл max}} / K_{\text{осл min}}$, равного 5,17; 5,35; 5,29; 5,73 для каждой модели соответственно (табл. 4). Такое изменение объясняется тем, что высокочастотные компоненты зашумленного сигнала проявляются в установившемся состоянии в большей степени, результатом чего является снижение подавления помехи, что приводит к увеличению $D_{\text{вых}}(\xi)$ и снижению $K_{\text{осл}}$ до единицы при $D_{\text{вх}}(\xi) = 0,1$ для каждой модели (рис. 5).

Таблица 4

Значения относительного коэффициента $K_{\text{осл отн}}^* = K_{\text{осл max}} / K_{\text{осл min}}$ для моделей M_1 , M_2 , M_3 , M_4 объекта управления при различных значениях дисперсии $D_{\text{вх}}(\xi)$

$D_{\text{вх}}(\xi)$	Модель объекта управления $M_i (i = 1 \dots 4)$ с передаточной функцией $W_{M_i}(p) (i = 1 \dots 4)$			
	$M_1(W_{M1}(p))$	$M_2(W_{M2}(p))$	$M_3(W_{M3}(p))$	$M_4(W_{M4}(p))$
	$K_{\text{осл}}$			
0,10	1,044	1,049	1,045	1,051
0,75	1,330	1,362	1,114	1,378
1,50	1,659	1,724	1,341	1,755
2,50	2,098	2,211	1,679	2,237
5,00	3,197	3,440	3,267	3,511
10,0	5,398	5,614	5,534	6,018

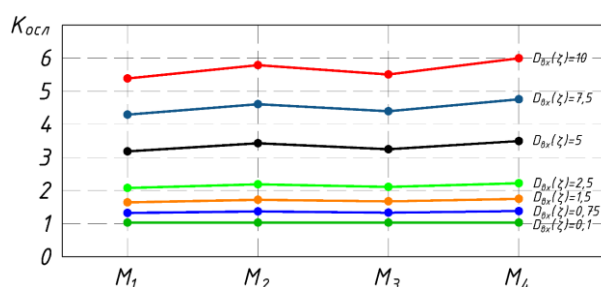


Рис. 5. Зависимость относительного коэффициента ослабления $K_{\text{осл}}$ для моделей M_1 , M_2 , M_3 , M_4 объекта управления при различных значениях дисперсии $D_{\text{вх}}(\xi)$

Из рис. 5 следует, что при изменении дисперсии на входе зависимость $K_{\text{осл}}$ идентична для исследуемых моделей интервального объекта, причем значение $K_{\text{осл}}$ находится в диапазоне от 1,044 до 5,398; от 1,049 до 5,614; от 1,045 до 5,534; от 1,051 до 6,018 для моделей M_1 , M_2 , M_3 , M_4 соответственно, а изменение относительного коэффициента ослабления $K_{\text{осл отн}}^*$ незначительно и уменьшается в $5,73/5,17 = 1,1$ раза. Величина $K_{\text{осл отн}}^*$ остается одинаковой как по оси ординат для каждой модели M_1 , M_2 , M_3 , M_4 объекта управления, так и по оси абсцисс для каждого значения $D_{\text{вх}}(\xi)$, поскольку $K_{\text{осл отн}}^*$ зависит от отношения абсолютных величин дисперсий, а не от их отклонений.

Анализ изменения $K_{\text{осл}}$ с увеличением дисперсии на входе $D_{\text{вх}}(\xi)$ позволяет отметить, что качество управления при ПИД-регуляторе с ДК, синтезированном при ступенчатом детерминированном воздействии, может снижаться. Причиной этого может быть то, что случайные помехи (шумы измерений) на входе, характеризуемые увеличением $D_{\text{вх}}(\xi)$, усиливаются производными ДК, а, следовательно, ослабление $D_{\text{вых}}(\xi)$ снижается. Эта особенность может проявляться при постоянной структуре САУ с регулятором и ДК для всех моделей объекта. Для уменьшения этого влияния на качество управления в ходе исследований установлено, что эффективным является подстройка параметров НС с помощью сигмоидных функций в обратных связях наблюдателя.

В ходе исследований установлено более значимое влияние коэффициента сигмоидной функции λ и коэффициента обратной связи l_1 на регулируемую переменную $Y(t)$ и оценку регулируемой переменной $\hat{Y}(t)$, что в условиях невозможности реализации скользящего режима является практически полезным. Для наглядности на рис. 6, 7 представлены графики регулируемой переменной $Y(t)$ и оценки регулируемой переменной $\hat{Y}(t)$ при дисперсии помехи $D(\xi) = 0,1$.

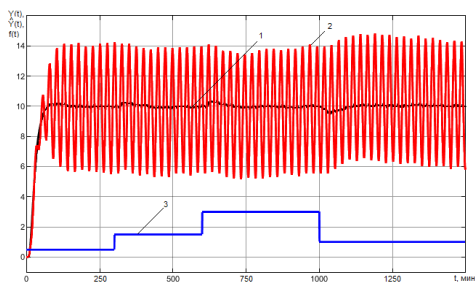


Рис. 6. График регулируемой переменной $Y(t)$ (кривая 1), оценки регулируемой переменной $\hat{Y}(t)$ (кривая 2), ступенчатого возмущения (кривая 3) при $l_1 = 1,0$, $l_2 = 0,0034$, $l_3 = 0,0638$, $\lambda = 1$

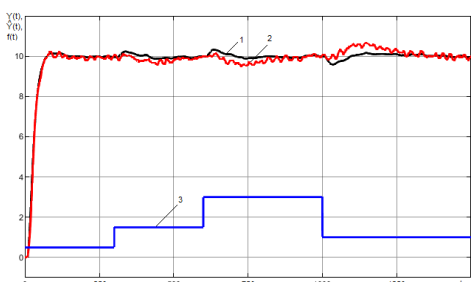


Рис. 7. График регулируемой переменной $Y(t)$ (кривая 1), оценки регулируемой переменной $\hat{Y}(t)$ (кривая 2), ступенчатого возмущения (кривая 3) при $l_1 = 0,5$, $l_2 = 0,0034$, $l_3 = 0,0638$, $\lambda = 0,5$

Из графиков на рис. 6, 7 видно, что при значении $l_1 = 1$ можно получить большую частоту переключений и амплитуду переменной на выходе НС, что дает возможность исключить большие динамические отклонения зашумленной регулируемой переменной $Y(t)$ на втором, третьем и четвертом участках (рис. 6). При уменьшении коэффициента l_1 в 2 раза ($l_1 = 0,5$) большие частота и колебания оценки $\hat{Y}(t)$ регулируемой переменной $Y(t)$ исключаются, но сохраняются приемлемые отклонения регулируемой переменной $Y(t)$ от заданного значения на втором, третьем и четвертом участках (рис. 7), что позволяет говорить о робастности системы к помехам в условиях ограниченного по амплитуде и частоте сигнала управления.

Для наглядности на рис. 8–11 показаны графики отклонений регулируемой переменной $Y(t)$ (парциального давления растворенного в биореакторе кислорода) и оценки регулируемой переменной $\hat{Y}(t)$ от заданного значения при изменении дисперсии помехи, а также графики изменения СКО регулируемой переменной $Y(t)$ и ее оценки $\hat{Y}(t)$ при изменении дисперсии.

Коэффициент ослабления в системе с моделью $W(p) = 0,375 / (219,04p^2 + 17,62p + 1) \cdot e^{-7,5p}$ при коэффициентах $l_1 = 0,1778$, $l_2 = 0,0034$, $l_3 = 0,0638$ составил $K_{осл} = 0,2436 / 0,1435 = 1,697$ (при $D_{вх}(\xi) = 0,25$); $K_{осл} = 0,43 / 0,18 = 2,388$ (при $D_{вх}(\xi) = 0,5$); $K_{осл} = 12,51 / 2,518 = 4,968$ (при $D_{вх}(\xi) = 10$).

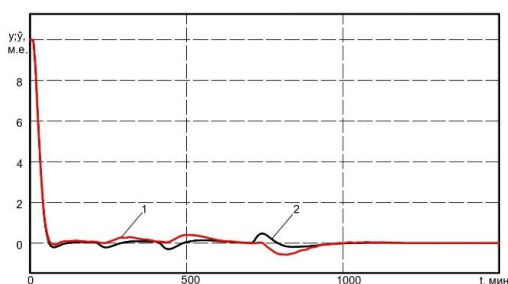


Рис. 8. Графики отклонений регулируемой переменной $Y(t)$ (кривая 1) и оценки регулируемой переменной $\hat{Y}(t)$ (кривая 2) от заданного значения при дисперсии $D_{вх}(\xi) = 0,1$

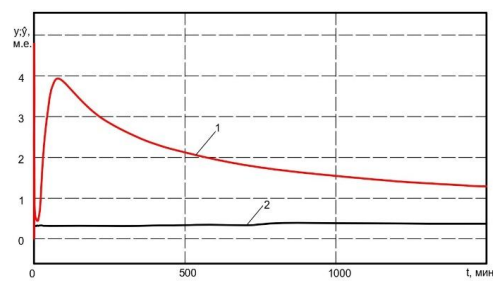


Рис. 9. Графики изменения СКО регулируемой переменной $Y(t)$ (кривая 1) и оценки регулируемой переменной $\hat{Y}(t)$ (кривая 2) при дисперсии $D_{вх}(\xi) = 0,1$

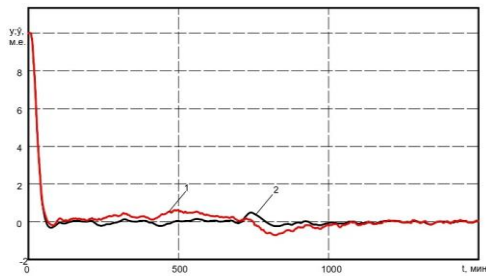


Рис. 10. Графики отклонений регулируемой переменной $Y(t)$ (кривая 1) и оценки регулируемой переменной $\hat{Y}(t)$ (кривая 2) от заданного значения при дисперсии $D_{ex}(\xi) = 10$

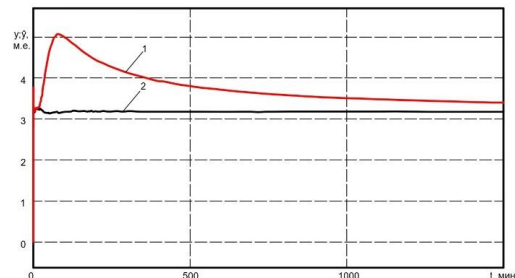


Рис. 11. Графики изменения СКО регулируемой переменной $Y(t)$ (кривая 1) и оценки регулируемой переменной $\hat{Y}(t)$ (кривая 2) при дисперсии $D_{ex}(\xi) = 10$

Результаты исследования систем. Оценку достижения параметрической робастности в рамках исследуемой структуры системы дает анализ переходных процессов, приведенных на рис. 3 и 4. Эти переходные процессы показывают зону робастности, в которой располагаются переходные процессы при промежуточных значениях $D_{bx}(\xi) = 0, 1; 1, 5$ и др. Анализ переходных процессов показывает, что для всех вариантов модели объекта с различным сочетанием параметров в установившемся режиме при постоянном задающем воздействии и при различных значениях возмущения $f(t)$ и дисперсии помехи $D_{bx}(\xi)$ обеспечивается робастность САУ по отношению к параметрической неопределенности параметров модели объекта и к действию возмущений и помех в канале измерения. САУ имеет хорошие демпфирующие свойства, которые незначительно зависят от внешнего возмущения, величины запаздывания и коэффициента передачи объекта. Ступенчатое изменение сигнала возмущения $f(t)$ в установившемся режиме приводит к приблизительно одинаковым результатам для моделей объектов $W_{M1}(p)$, $W_{M2}(p)$, $W_{M3}(p)$, $W_{M4}(p)$. Следует отметить, что в этом случае отличия сравнительных показателей оказываются меньшими, чем при отработке одновременного действия возмущений в переходном и в установившемся режиме. Это объясняется тем, что при ступенчатом изменении сигнала возмущающего воздействия $f(t)$ на входе объекта в меньшей степени проявляются высокочастотные составляющие регулируемой переменной (их возможное усиление производными компенсирующего устройства), чем при изменении задающего воздействия, для которого обычно проводится синтез регулятора.

Результаты моделирования подтверждают робастность и эффективность разработанной методики синтеза.

Заключение. Показана возможность практической реализации структуры системы с ДК инерционности объекта управления, реализуемой на основе оценок переменных состояний, получаемых с помощью НС с обратными связями сигмоидного типа. Устойчивое поведение системы с ДК и НС получено для всех моделей объекта при различных значениях дисперсии помехи. Следовательно, синтез робастной системы управления объектом с типовым регулятором и ДК, реализуемым на основе оценок производных с помощью НС, как показали исследования, оказывается возможным при больших изменениях параметров объекта и помехи в измеряемом сигнале. Следует также отметить, что использование сигмоидных функций в корректирующих обратных связях наблюдателя способствует повышению темпа оценивания и уменьшению ошибок, связанных с неточностью математической модели объекта.

Проведенными исследованиями САУ на примере промышленного биообъекта было доказано, что возможна практически реализуемая структура системы с постоянными параметрами типового ПИД-регулятора, ДК и НС, обеспечивающая робастное управление нестационарным объектом, описываемым интервальной моделью с постоянными параметрами. Установлено, что усиление помехи производными, реализуемыми в ДК с по-

мощью оценок производных, получаемых в НС, не существенно. Можно считать, что измерения осуществляются с допустимой ошибкой и полученные измерения и оценки можно считать приемлемыми.

Следует отметить, что в случае действия только возмущения в установившемся состоянии отличия сравнительных показателей оказываются меньшими, чем при отработке одновременного воздействия обоих возмущений по каналу задания и возмущения в процессе вывода на установившийся режим. Это объясняется тем, что при ступенчатом изменении только сигнала возмущающего воздействия $f(t)$ на входе объекта при постоянном задании в меньшей степени проявляются высокочастотные составляющие регулируемой переменной, а, следовательно, меньшее возможное их усиление производными ДК, чем при изменении задающего воздействия.

Проведенными исследованиями установлено, что структура САУ с типовым ПИД-регулятором в сочетании с последовательным ДК инерционности объекта и НС с корректирующими обратными связями на основе сигма-функций обеспечивает простоту методики синтеза САУ, робастной к изменениям параметров объекта и действию неизмеряемых возмущений и неконтролируемых помех.

Работа выполнена в Кубанском государственном технологическом университете, Краснодар.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Лубенцов В.Ф., Лужевский Н.О., Лубенцова Е.В. Анализ подходов к синтезу систем управления объектами с запаздыванием с использованием наблюдателей состояния // Электронный сетевой политематический журнал «Научные труды КубГТУ». – 2024. – № 4. – С. 122-132.
2. Пухов Г.Е., Жук К.Д. Синтез многосвязных систем управления по методу обратных операторов. – Киев: Наукова думка, 1966. – 218 с.
3. Изерман Р. Цифровые системы управления: пер. с англ. – М.: Мир, 1984. – 541 с.
4. Лужевский Н.О., Шахрай Е.А., Лубенцов В.Ф., Лубенцова Е.В. Анализ и синтез робастных регуляторов с типовыми законами регулирования для квазистационарных объектов с запаздыванием // Приборы и системы. Управление, контроль, диагностика. – 2024. – № 5. – С. 24-31.
5. Краснова С.А., Уткин А.В. Сигма-функция в задачах синтеза наблюдателей состояний и возмущений // Проблемы управления. – 2015. – № 5. – С. 27-36.
6. Потапенко Е.М., Казурова А.Е. Основы теории автоматического управления. – Запорожье: ЗНТУ, 2007. – 162 с.
7. Денисенко В. ПИД-регуляторы: принципы построения и модификации. Ч. 1 // Современные технологии автоматизации. – 2006. – № 4. – С. 66-74.
8. Тутов И.А. Система автоматического регулирования давления внутрипромысловой газораспределительной сети // Актуальные проблемы гуманитарных и естественных наук. Ч. 1. – 2013. – № 11 (58). – С. 118-121.
9. Шубладзе А.М., Попадько В.Е., Якушева А.А., Кузнецов С.И. Исследование оптимальных по степени устойчивости решений при ПИД управлении. Ч. 1 // Управление большими системами: сборник трудов. – 2008. – Вып. 22. – С. 86-100.
10. Александров А.Г., Паленов М.В. Состояние и перспективы развития адаптивных ПИД-регуляторов // Автоматика и телемеханика. – 2014. – № 2. – С. 16-30.
11. Колпиков Ю.Г., Лубенцов В.Ф., Опришко А.А., Белова Д.А., Шубладзе А.М. Об особенностях использования адаптивного ПИ-регулятора с активной идентификацией // Приборы и системы управления. – 1989. – № 4. – С. 26-27.
12. Патент № 2234116 С1 Российская Федерация, МПК G05B 11/36. Пропорционально-интегрально-дифференциальный регулятор: № 2002135205/09; заявл. 24.12.2002; опубл. 10.08.2004 / Е.В. Лубенцова; заявитель Северо-Кавказский государственный технический университет.
13. Гайдук А.Р. Теория и методы аналитического синтеза систем автоматического управления (полиномиальный подход). – М.: Физмалит. 2012. – 360 с.
14. Дудников Е.Г. Автоматическое управление в химической промышленности: учебник для вузов / под ред. Е.Г. Дудникова. – М.: Химия, 1987. – 368 с.
15. Гайворонский С.А., Езангина Т.А., Соболев А.В. Синтез систем управления максимальной робастной степени устойчивости на основе вершинных критических корневых диаграмм // Мехатроника, автоматизация, управление. – 2023. – № 24 (10). – С. 519-525.

16. Тычинин А.В. Структурно-параметрический синтез системы граничного управления объектом с распределенными параметрами методами обратной динамики: специальность 05.13.01 "Системный анализ, управление и обработка информации (по отраслям)": автореф. дисс. ... канд. тех. наук / А.В.Тычинин. – Самара, 2008. – 22 с.
17. Филимонов А.Б., Филимонов Н.Б. К вопросу синтеза систем многорежимного регулирования // Journal of Advanced Research in Technical Science. – 2020. – № 18. – С. 41-49.
18. Ремизова О.А., Сыроквашиин В.В., Фокин А.Л. Синтез робастных систем управления с типовыми регуляторами // Известия высших учебных заведений. Приборостроение. – 2015. – Т. 58, № 12. – С. 966-972.
19. Тенденции развития методов синтеза систем автоматического управления (САУ). – URL: <https://gigabaza.ru/doc/112825-pall.html> (дата обращения: 25.10.2025).
20. Фокин А.Л. Синтез робастных систем управления технологическими процессами с типовыми регуляторами // Известия Санкт-Петербургского государственного технологического института (технического университета). – 2014. – № 27 (53). – С. 101-106.
21. Лубенцова Е.В., Лубенцов В.Ф. Построение оптимальных по быстродействию систем стабилизации // Математические методы в технике и технологиях: ММТТ-2000: Сб. трудов 13 Междунар. науч. конф.: В 7 т. Т. 6. / Редкол.: Балакирев В.С. и др. – СПб.: С.-Петербург. гос. технол. ин-т (техн. ун-т), 2000. – С. 272-273.
22. Антипов А.С., Краснова С.А., Уткин В.А. Синтез инвариантных нелинейных одноканальных систем слежения с сигмоидальными обратными связями с обеспечением заданной точности слежения // Автоматика и телемеханика. – 2022. – № 1. – С. 40-66.
23. Gouze J.-L., Rapaport A., Hady-Sadok Z. Interval observers for uncertain biological systems // Ecological modelling. – Issue 133 (1). – P. 45-56.
24. Bernard O., Gouze J. L. Closed loop observers bundle for uncertain biotechnological models // Journal of Process Control. – October 2004. – Volume 14. – Issue 7. – P. 765-774.
25. Alessandri A., Baglietto M., Battistelli G. Design of state estimators for uncertain linear systems using quadratic boundedness // Automatica. – 2006. – Volume 42. – P. 497-502.
26. Харьковская Т.А., Кремлев А.С., Сабирова Д.М., Ефимов Д.В., Раисси Т. Интервальный наблюдатель для модели биологического реактора // Научно-технический вестник информационных технологий, механики и оптики. – 2014. – № 3 (91). – С. 39-45.
27. Пьявченко Т.А., Финаев В.И. Автоматизированные информационно-управляющие системы. – Таганрог: Изд-во ТРТУ, 2007. – 271 с.

REFERENCES

1. Lubentsov V.F., Luzhevskiy N.O., Lubentsova E.V. Analiz podkhodov k sintezu sistem upravleniya ob'yektami s zapazdyvaniyem s ispol'zovaniyem nablyudateley sostoyaniya [Analysis of approaches to the synthesis of control systems for objects with delay using state observers]. *Elektronnyy setevoy politematischeskiy zhurnal «Nauchnyye trudy KubGTU»* [Electronic network polythematic journal «Scientific works of KubSTU»], 2024, No 4, pp. 122-132.
2. Pukhov G.E., Zhuk K.D. Sintez mnogovyaznykh sistem upravleniya po metodu obratnykh operatorov [Synthesis of Multivariate Control Systems Using the Inverse Operator Method]. Kiev: Naukova dumka, 1966, 218 p.
3. Izerman R. Tsifrovyye sistemy upravleniya: transl. from Engl. [Digital Control Systems: Translated from English]. Moscow: Mir, 1984, 541 p.
4. Luzhevskiy N.O., Shakhrai E.A., Lubentsov V.F., Lubentsova E.V. Analiz i sintez robastnykh regulyatorov s tipovymi zakonami regulirovaniya dlya kvazistatsionarnykh ob'yektov s zapazdyvaniyem [Analysis and synthesis of robust regulators with standard control laws for quasi-stationary objects with delay], *Pribory i sistemy. Upravleniye, kontrol', diagnostika* [Instruments and Systems. Control, Monitoring, Diagnostics], 2024, No. 5, pp. 24-31.
5. Krasnova S.A., Utkin A.V. Sigma-funktsiya v zadachakh sinteza nablyudateley sostoyaniy i vozmushcheniy [Sigma Function in Problems of Synthesis of State and Disturbance Observers], *Problemy upravleniya* [Problems of Control], 2015, No. 5, pp. 27-36.
6. Potapenko E.M., Kazurova A.E. *Osnovy teorii avtomaticheskogo upravleniya* [Fundamentals of Automatic Control Theory]. Zaporozh'e: ZNTU, 2007, 162 p.
7. Denisenko V. PID-regulyatory: printsipy postroyeniya i modifikatsii. Chast' 1 [PID Controllers: Principles of Design and Modification. Part 1], *Sovremennyye tekhnologii avtomatizatsii* [Modern Automation Technologies], 2006, No. 4, pp. 66-74.
8. Tutov I.A. Sistema avtomaticheskogo regulirovaniya davleniya vnutripromyslovoy gazoraspredeletel'noy seti [Automatic pressure control system for an in-field gas distribution network]. *Aktual'nyye problemy gumanitarnyy i yestestvennykh nauk. Chast' 1* [Actual problems of humanitarian and natural sciences, Part 1], 2013, No. 11 (58), pp. 118-121.

9. Shubladze A.M., Popad'ko V.E., Yakusheva A.A., Kuznetsov S.I. Issledovaniye optimal'nykh po stepeni ustoychivosti resheniy pri PID upravlenii. Chast' 1 [Study of optimal solutions in terms of stability under PID control. Part 1], *Upravleniye bol'shimi sistemami: sbornik trudov* [Control of large systems: collected papers], 2008, Issue. 22, pp. 86-100.
10. Aleksandrov A.G., Palenov M.V. Sostoyaniye i perspektivy razvitiya adaptivnykh PID-regulyatorov [Status and development prospects of adaptive PID controllers], *Avtomatika i telemekhanika* [Automation and Telemechanics], 2014, No 2, pp. 16-30.
11. Kolpikov Yu.G., Lubentsov V.F., Oprishko A.A., Belova D.A., Shubladze A.M. Ob osobennostyakh ispol'zovaniya adaptivnogo PI-regulyatora s aktivnoy identifikatsiyey [On the Features of Using an Adaptive PI Controller with Active Identification], *Pribory i sistemy upravleniya* [Control Instruments and Systems], 1989, No. 4, pp. 26-27.
12. Lubentsova E.V. Proportsional'no-integral'no-differentsial'nyy regulyator [Proportional-integral-differential controller]: Patent RF, no 2234116 C1, 2004.
13. Gayduk A.R. Teoriya i metody analiticheskogo sinteza sistem avtomaticheskogo upravleniya (polinomial'nyy podkhod) [Theory and Methods of Analytical Synthesis of Automatic Control Systems (Polynomial Approach)], Moscow: Fizmatlit, 2012, 360 p.
14. Dudnikov E.G. *Avtomaticheskoye upravleniye v khimicheskoy promyshlennosti* [Automatic Control in the Chemical Industry], Moscow: Khimiya, 1987, 368 p.
15. Gayvoronskiy S.A., Yezangina T.A., Sobol' A.V. Sintez sistem upravleniya maksimal'noy robustnoy stepeni ustoychivosti na osnove vershinnykh kriticheskikh kornevykh diagramm [Synthesis of Maximum Robust Stability Control Systems Based on Vertex Critical Root Diagrams], *Mekhatronika, avtomatizatsiya, upravleniye* [Mechatronics, Automation, Control], 2023, No. 24 (10), pp. 519-525.
16. Tychinin A.V. Strukturno-parametricheskii sintez sistemy granichnogo upravleniya ob'ektom s raspredelennymi parametrami metodami obratnoy dinamiki: spetsial'nost' 05.13.01 "Sistemnyy analiz, upravlenie i obrabotka informatsii (po otraslyam)": avtoref. diss. ... kand. tekhn. nauk [Structural-parametric synthesis of a boundary control system for an object with distributed parameters using inverse dynamics methods: abstract cand. of eng. sc. diss.]. Samara, 2008, 22 p.
17. Filimonov A.B., Filimonov N.B. K voprosu sinteza sistem mnogorezhimnogo regulirovaniya [On the issue of synthesis of multimode control systems], *Journal of Advanced Research in Technical Science* [Journal of Advanced Research in Technical Science], 2020, No. 18, pp. 41-49.
18. Remizova O.A., Syrovkashin V.V., Fokin A.L. Sintez robustnykh sistem upravleniya s tipovymi regulyatorami [Synthesis of robust control systems with standard controllers], *Izvestiya vysshihkh uchebnykh zavedeniy. Priborostroyeniye* [News of higher educational institutions. Instrument engineering], 2015, Vol. 58, No. 12, pp. 966-972.
19. Tendentsii razvitiya metodov sinteza sistem avtomaticheskogo upravleniya (SAU) [Trends in the Development of Methods for Synthesizing Automatic Control Systems (ACS). Available at: <https://gigabaza.ru/doc/112825-pall.html> (accessed 25 October 2025).
20. Fokin A.L. Sintez robustnykh sistem upravleniya tekhnologicheskimi protsessami s tipovymi regulyatorami [Synthesis of Robust Process Control Systems with Standard Controllers], *Izvestiya Sankt-Peterburgskogo gosudarstvennogo tekhnologicheskogo instituta (tekhnicheskogo universiteta)* [Bulletin of the St. Petersburg State Technological Institute (Technical University)], 2014, No. 27 (53), pp. 101-106.
21. Lubentsova E.V., Lubentsov V.F. Postroyeniye optimal'nykh po bystrodeystviyu sistem stabilizatsii [Construction of stabilization systems with optimal response time], *Matematicheskiye metody v tekhnike i tekhnologiyakh: MMTT-2000: Sb. tr. 13 Mezhdunar. nauch. konf.* [Mathematical methods in engineering and technology: MMTT-2000: Collection of works of the 13th International scientific conference]: In 7 vol. Vol. 6. St. Petersburg: St. Petersburg State Technological Institute (Technical University), 2000, pp. 272-273.
22. Antipov A.S., Krasnova S.A., Utkin V.A. Sintez invariantnykh nelineynykh odnokanal'nykh sistem slezheniya s sigmoidal'nymi obratnymi svyazyami s obespecheniyem zadannoy tochnosti slezheniya [Synthesis of invariant nonlinear single-channel tracking systems with sigmoid feedbacks ensuring a specified tracking accuracy], *Avtomatika i telemekhanika* [Automation and Telemechanics], 2022, No 1, pp. 40-66.
23. Gouze J.-L., Rapaport A., Hady-Sadok Z. Interval observers for uncertain biological systems, *Ecological modeling*, Issue 133 (1), pp. 45-56.
24. Bernard O., Gouze J. L. Closed loop observers bundle for uncertain biotechnological models, *Journal of Process Control*, October 2004, Vol. 14, Issue 7, pp. 765-774.
25. Alessandri A., Baglietto M., Battistelli G. Design of state estimators for uncertain linear systems using quadratic boundedness, *Automatica*, 2006, Vol. 42, pp. 497-502.

26. Khar'kovskaya T.A., Kremlev A.S., Sabirova D.M., Efimov D.V., Raissi T. Interval'nyy nablyudatel' dlya modeli biologicheskogo reaktora [Interval observer for a biological reactor model], *Nauchno-tekhnicheskiiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki*. [Scientific and Technical Journal of Information Technologies, Mechanics and Optics], 2014, No 3 (91), pp. 39-45.
27. Pyavchenko T.A., Finayev V.I. *Avtomatizirovannyye informatsionno-upravlyayushchiye sistemy* [Automated information and control systems]. Taganrog: Izd-vo TRTU, 2007, 271 p.

Лужевский Никита Олегович – Кубанский государственный технологический университет; e-mail: nikitaluzhev97@gmail.com; г. Краснодар, Россия; тел.: 89288833458; ассистент кафедры автоматизации производственных процессов.

Лубенцов Валерий Федорович – Кубанский государственный технологический университет; e-mail: vf.lubentsov@yandex.ru; г. Краснодар, Россия; тел.: 89614440061; д.т.н.; доцент; профессор кафедры автоматизации производственных процессов.

Лубенцова Елена Валерьевна – Кубанский государственный технологический университет; e-mail: lubentsovaev@mail.ru; г. Краснодар, Россия; тел.: 89034182575; д.т.н.; доцент; профессор кафедры автоматизации производственных процессов.

Luzhevsky Nikita Olegovich – Kuban State Technological University; e-mail: nikitaluzhev97@gmail.com; Krasnodar, Russia; phone: +79288833458; assistant professor, Department of Automation of Industrial Processes.

Lubentsov Valery Fedorovich – Kuban State Technological University; e-mail: vf.lubentsov@yandex.ru; Krasnodar, Russia; phone: +79614440061; dr. of eng. sc.; associate professor; professor, Department of Automation of Industrial Processes.

Lubentsova Elena Valeryevna – Kuban State Technological University; e-mail: lubentsovaev@mail.ru; Krasnodar, Russia; phone: +79034182575; dr. of eng. sc.; associate professor; professor, Department of Automation of Industrial Processes.

УДК 004.81 + 681.5.015

DOI 10.18522/2311-3103-2026-1-164-179

Д.Н. Богачева, О.В. Лукинова, А.А. Саломатин

СЕМАНТИЧЕСКАЯ МОДЕЛЬ ВЛИЯНИЯ ФАКТОРОВ БЕЗОПАСНОСТИ ДЛЯ ОЦЕНКИ КИБЕРУСТОЙЧИВОСТИ ИНФОРМАЦИОННЫХ СИСТЕМ

В современных условиях обеспечение киберустойчивости информационно-телекоммуникационных систем предприятий становится приоритетной задачей, требующей учета сложного взаимодействия множества факторов. Данная работа посвящена исследованию устойчивости информационно-телекоммуникационных систем с точки зрения процессно-ориентированного подхода, который фокусируется на сохранении непрерывного и безопасного функционирования предприятия при возникновении киберинцидентов. При этом одна из существующих сегодня проблем заключается в отсутствии средств системного анализа факторов безопасности, их влияния на общую защищенность бизнес-процессов. Цель исследования заключается в разработке формализованного инструмента, позволяющего прогнозировать уровень устойчивости с учетом функциональных зависимостей между факторами. В качестве основного метода исследования выступает концептуальное семантическое моделирование, позволяющее формализовать причинно-следственные связи между элементами системы. Научная новизна заключается в разработке модели, фиксирующей взаимное влияние угроз и мер противодействия, что позволяет прогнозировать уровень устойчивости информационно-телекоммуникационных систем уже на этапе проектирования. На примере метода оценки корректности параметров оконечных устройств автоматизированной системы управления технологическим процессом продемонстрировано влияние различных факторов на эффективность механизмов защиты нижнего уровня системы. Результаты, представленные в работе, позволяют сократить противоречие между научным и практическим применением процессно-ориентированного подхода, а также могут быть использованы как теоретическая основа для разработки программных инструментов в контексте принятия решений по выбору мер защиты бизнеса на основе баланса между возможным ущербом, ожидаемыми уровнями киберустойчивости, рисков и безопасности в течение всего жизненного цикла информационной системы.

Процессно-ориентированный подход; информационная система; киберустойчивость; семантическая модель факторов безопасности; датчик.

D.N. Bogacheva, O.V. Lukinova, A.A. Salomatin

SEMANTIC MODEL OF SECURITY FACTORS INFLUENCE FOR EVALUATING CYBER RESILIENCE OF INFORMATION SYSTEMS

In modern conditions, ensuring the cyber resilience of enterprise information and telecommunication systems is becoming a priority task that requires accounting for the complex interaction of numerous factors. This paper investigates resilience of information and telecommunication systems through the lens of a process-oriented approach, which focuses on maintaining continuous and secure enterprise operations in the event of cyber incidents. However, one of the current problems lies in the lack of tools for systematic analysis of security factors and their impact on the overall protection of business processes. The aim of the study is to develop a formalized tool for predicting resilience levels, taking into account functional dependencies between factors. The primary research method is conceptual semantic modeling, which enables the formalization of cause-and-effect relationships between system elements. The scientific novelty consists in the development of a model that captures the mutual influence of threats and counter-measures, allowing for the prediction of information and telecommunication system resilience levels as early as the design stage. Using the method for evaluating the correctness of Industrial Control Systems endpoint parameters as an example, the influence of various factors on the effectiveness of the system's lower-level protection mechanisms is demonstrated. The results presented in the paper help reduce the gap between the theoretical and practical application of the process-oriented approach. They can also serve as a theoretical foundation for developing software tools to support decision-making in selecting business protection measures, based on a balance between potential damage, expected levels of cyber resilience, risks, and security throughout the entire lifecycle.

Process-oriented approach; information system; cyber resilience; semantic model of security factors; sensor.

Введение. Одной из наиболее важных особенностей развития современных информационных систем (ИС) следует считать тенденцию к интеграции. Речь идет о функциональной интеграции, интеграции неоднородных информационных ресурсов, интеграции разных способов представления информации для пользователей информационных систем, интеграции информационных и/или вычислительных систем с телекоммуникационными системами, т.е. возникновению класса интегрированных информационно-телекоммуникационных систем (ИТКС), в которых объединяются функции ИС и систем передачи данных.

К классу ИТКС относятся различные распределенные корпоративные информационные системы (КИС):

- ♦ системы, обслуживающие предприятия различных отраслей экономики, финансово-кредитной сферы, системы военного назначения и др.;
- ♦ государственные ИС, созданные для нужд правоохранительных органов, электронного правительства, осуществления государственных услуг и т.п.;
- ♦ глобальные информационные сети, в которых традиционные услуги связи и передачи данных дополняются услугами информационного обслуживания массовых пользователей.

Применение Web-технологий в ИТКС обеспечивает развитие таких областей, как электронный бизнес (оптовая, вплоть до виртуальных бирж, и розничная торговля – Интернет-супермаркеты; взаимодействие промышленных предприятий с потребителями и поставщиками – *business-to-business*), электронные библиотеки, дистанционное обучение (вплоть до виртуальных университетов).

ИТКС представляют собой наиболее сложный класс современных ИС с точки зрения методов и средств их создания, сопровождения и развития, инструментов проектирования и программирования приложений, средств представления информационных ресурсов и доступа к ним.

Очевидно, что к ИТКС предъявляются повышенные требования, в том числе и к вопросам устойчивого функционирования по отношению к внешним негативным воздействиям нарушителя, например, несанкционированному доступу к данным при их

хранении, обработке и передаче по каналам связи или организации распределенных DOS-атак. Анализ источников позволяет сделать вывод о том, что устойчивость ИС рассматривается в следующих взаимосвязанных аспектах.

1. Структурно-топологическая устойчивость сетевого слоя ИТКС. В работе [1] приведено определение инфраструктурной устойчивости: «способность инфраструктуры при возмущении системы оставаться на заданном качественном уровне на фоне высокого уровня инфраструктурного деструктивизма». Под инфраструктурным деструктивизмом здесь понимается неспособность какого-либо компонента ИТКС выполнять свои функции в полном объеме.

2. Функциональная устойчивость. В руководстве по оценке мер безопасности в федеральных информационных системах [2], опубликованном Национальным институтом стандартов и технологий США (NIST), в исследованиях [3–7] вводится понятие функциональной устойчивости как способность сохранения и/или восстановления возможности выполнения возложенных на систему функций при различных условиях существования, в том числе и при деструктивных воздействиях на ее элементы.

3. Киберустойчивость. В работе [2] киберустойчивость определяется как способность предвидеть, выдерживать, восстанавливаться и адаптироваться к неблагоприятным условиям, атакам или компрометациям ИТ-систем. Всемирный экономический форум (WEF) определяет киберустойчивость как способность организации минимизировать влияние существенных киберинцидентов на ее основные бизнес-цели и задачи. Здесь прослеживается связь с методологией, описанной в [8–10], где речь идет о том, что систему защиты ИС необходимо планировать, проектировать и эксплуатировать исходя из обеспечения безопасности бизнес-процессов организации. Среди стратегий обеспечения основ киберустойчивости эксперты WEF называют управление рисками. Риск – это потенциальная опасность нанесения ущерба организации в результате реализации некоторой угрозы с использованием уязвимостей актива или группы активов. Определяется как функция от вероятности реализации угрозы, использующей уязвимости, и оценки величины ущерба, как последствия реализации угрозы [11, 12]. Управление рисками включает: определение профиля рисков; назначение явных владельцев конкретных рисков и определение ответственности в случае их наступления; планирование и внедрение мер снижения и смягчения рисков; соблюдение регуляторных требований [13]. В работе [14] также рассматривается связь между устойчивостью и рисками негативных воздействий на компоненты ИТКС. Для каждого компонента экспертами задаются критерии – свойства безопасности (конфиденциальность, целостность и доступность), а также коэффициенты важности, степень уязвимости актива и степень реализации угрозы по каждому критерию. Эти параметры авторы используют для расчета рисков нарушения свойств безопасности, посредством которых, в свою очередь, оценивается совокупная мера риска всей ИТКС. В работе [15] авторы подчеркивают, что наиболее важной составляющей в обеспечении киберустойчивости и, как следствие, общей устойчивости, является именно оценка рисков, т.к. именно регулярный анализ рисков помогает разработать эффективные стратегии защиты, а также минимизировать вероятность различных кибератак.

Кроме того, повсеместное развитие цифровой экономики требует от бизнес-структур непрерывного функционирования, что, в свою очередь, реализуемо только при внедрении в жизненный цикл ИТКС принципов процессно-ориентированного подхода [16], которые переводят фокус внимания защиты с информационной системы на бизнес-организацию предприятия.

Тогда очевидно, что возникает проблема в разработке таких инструментов, которые позволили бы еще при создании ИС прогнозировать уровень ее устойчивости к воздействиям различных инцидентов, инициируемых злоумышленниками.

Для решения указанной задачи в данной работе предлагается семантическая модель, предназначенная для оценки влияния факторов безопасности на риски и киберустойчивость при проектировании или эксплуатации ИТКС. При этом модель учитывает расширенный (по сравнению с классическим) набор параметров безопасности, отражает принципы процессно-ориентированного подхода, включает концепт целевой функции, содер-

жит контуры принятия решений в разных парадигмах управления безопасностью и может служить формализованной основой для создания программных инструментов указанного подхода.

Постановка задачи. На сегодняшний день существует два основных подхода к проектированию систем обеспечения безопасности ИС. Первый – традиционный, классический, основан на удовлетворении требований нормативно-правовых документов [17–19], которые в качестве объекта защиты определяют саму целевую ИС, а также фокусируются на барьерных мерах по защите периметра ИС за счет встроенных и наложенных средств защиты (СЗ). Основная направленность этого подхода – противодействие предполагаемым угрозам в целях предотвращения несанкционированного доступа, сохранения целостности, обеспечения доступности данных, хранящихся и обрабатываемых в системе. Поэтому главным требованием регуляторов информационной безопасности (ИБ) является наличие для защищаемой системы модели угроз, т.е. моделей нарушителя, уязвимостей, а также потенциально возможных атак, противовесом которых являются механизмы и средства защиты.

Однако в современных условиях глобального интернета, когда размывается само понятие «защищаемый периметр» вследствие использования облачных сервисов; когда резко возросло количество кибератак и технических возможностей нарушителей для их осуществления; когда усложнились архитектуры систем; когда многократно возросла потребность в оценке влияния киберрисков на бизнес-цели, возникла новая парадигма – процессно-ориентированный подход. Первыми публикациями в этой области, в которых изложены базовые принципы новой методологии, можно считать [9, 10]. Эти принципы апеллируют к бизнес-процессному подходу, стандартам *ITIL* [20] и заключаются в том, что, в отличие от традиционной парадигмы, требования к безопасности информационной инфраструктуры предприятия необходимо выстраивать исходя из безопасного функционирования бизнес-процессов организации, их информационных потоков и функций. Следовательно, объектом защиты становятся бизнес-процессы компании, а не информационная система, которая в соответствии с *ITIL*, является лишь способом реализации бизнес-процесса, его ресурсом.

Таким образом, процессно-ориентированный подход представляет собой методологию, ориентированную на обеспечение безопасной и непрерывной деятельности бизнеса, методически основываясь на том, что свойства безопасности (конфиденциальность (*K*), целостность (*C*), доступность (*D*)) становятся атрибутами бизнес-процесса и, одновременно, требованиями безопасности, предъявляемыми к защите приложений и платформы информационной системы. Другими словами, вектор $\vec{KS} = \{K, C, D\}$ будет являться целевым при проектировании системы безопасности ИТКС. Тогда очевидно, что риски ИБ – это лишь один из подчиненных аспектов управления бизнес-рисками предприятия и, если они становятся для предприятия актуальными, то ИБ-защита выступает способом минимизации общих рисков. Международные и отечественные стандарты и методологии по управлению рисками [16, 21], начиная с десятих годов 21 века, также фиксируют основные положения процессно-ориентированного подхода. Так, стандарт NIST SP 800-39 «Managing Information Security Risk» [16] прямо описывает трехуровневую модель управления рисками: организация, бизнес-процессы, информационные системы, которая предписывает на уровне ИС реализовать все требования по управлению рисками ИБ, которые были приняты на бизнес-уровне. Кроме того, вводится понятие устойчивости ИС как показатель жизнеспособности бизнес-функций компании. Отечественный ГОСТ Р ИСО/МЭК 31010-2011 «Менеджмент риска. Методы оценки риска» [21] апеллирует к 607-П ЦБ РФ «О требованиях к порядку обеспечения бесперебойности функционирования платежной системы, показателям бесперебойности функционирования платежной системы и методикам анализа рисков в платежной системе, включая профили рисков». Методология *Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE)* используется для моделирования и оценки рисков как информационных систем, так и критических бизнес-активов компании. В стандарте ISO/IEC 27005:2018 [22] декларируется тот факт, что оценка рисков должна осуществляться с учетом последствий рисков для бизнеса.

Однако следует заметить, что, несмотря на наличие документов и публикаций, декларирующих положения прогрессивного подхода, в области практического использования указанных стандартов существуют следующие проблемы. Во-первых, нормативные документы носят исключительно рекомендательный характер. Во-вторых, несмотря на достаточно развитые теоретические и нормативные основания процессной методологии, наличие средств по моделированию угроз, анализу уязвимостей программного обеспечения ИС и пр., на практике существует потребность в таких моделях, методах и инструментах, которые воплощали бы процессно-ориентированную идеологию.

Таким образом, целью данного исследования является разработка модели, позволяющей оценивать взаимовлияние различных факторов ИБ при выборе механизмов и средств защиты от воздействия киберинцидентов, минимизирующих риски и повышающих устойчивое функционирование как ИТКС, так бизнес-процессов.

Для достижения указанной цели в ходе исследования необходимо выполнить следующие задачи:

1. Расширить, систематизировать и формализовать перечень ключевых факторов ИБ, влияющих на киберустойчивость и позволяющих проводить более детализированный анализ;
2. Разработать концептуальную семантическую модель, устанавливающую причинно-следственные связи между факторами безопасности. На основе модели выделить и проанализировать управленческие контуры, соответствующие различным парадигмам обеспечения безопасности (классической, ориентированной на риски, процессной и комплексной).
3. Показать, как экземпляр концепта механизмов защиты, реализованный на основе метода оценки корректности параметров оконечных устройств (датчиков) [23, 24], и расширяющий номенклатуру существующих средств, может быть использован при оценке анализа защищенности автоматизированных систем управления технологическим процессом (АСУ ТП).

Решение этих задач позволит создать формализованный аппарат для комплексного анализа факторов киберустойчивости, связывающий технические средства защиты с бизнес-требованиями и управлением рисками, который может быть реализован в качестве ядра системы поддержки принятия решений для выбора рационального набора защитных мер вдоль всего жизненного цикла системы.

Семантическая модель влияния факторов информационной безопасности. Для решения поставленной задачи в работе был выбран аппарат семантического моделирования, который широко применяется в области информационной безопасности. Наиболее распространенными среди них являются следующие средства.

1. Онтологические модели.

Онтологические модели применяются для систематизации основных понятий предметной области, позволяя связать виды угроз, их взаимовлияния друг на друга и на остальные концепты кибербезопасности [25].

Примеры онтологий в области кибербезопасности приведены в работах [26, 27].

Онтологические модели обычно описывают с использованием формальных языков разметки *Web Ontology Language (OWL)* и *Resource Description Framework (RDF)* [28].

OWL позволяет создавать четкие и однозначные структуры данных, удобные для автоматической обработки и анализа. Основываясь на логике первого порядка, *OWL* способен представлять сложные связи и ограничения между классами и свойствами, что делает его незаменимым инструментом для задач, требующих строгой формализации данных. Его ключевые преимущества включают высокую выразительную мощь, возможность логического вывода и поддержку семантического анализа данных. Однако сложность создания больших онтологий и высокие требования к производительности остаются основными препятствиями для массового внедрения.

RDF организует данные в виде триплетов «субъект-предикат-объект», что позволяет описывать любые сущности и их свойства в унифицированном формате. *RDF* подходит для описания разнообразной информации, совмещается с другими стандартами, способен легко внедряться в существующее программное обеспечение за счет подключае-

мых модулей и библиотек для различных языков программирования. Однако у *RDF* есть и ограничения, связанные с относительно низкой производительностью при работе с большими объёмами данных и недостатком встроенных средств проверки целостности.

2. Когнитивные карты.

Когнитивные карты, как показывает анализ, в ИБ зачастую применяются для моделирования угроз и оценки рисков. Они показывают потенциальные направления угроз, включая зоны наибольшего риска, пути движения злоумышленников и потенциальные последствия успешных атак. Они помогают организациям наглядно оценить свои слабые места и спланировать эффективные меры защиты.

Среди карт угроз распространены нечеткие когнитивные карты (НKK). Они являются удобными в использовании, имеют возможность учитывать неопределённость и нечеткость в процессе оценки рисков, способны изменяться и адаптироваться в соответствии с новыми данными и требованиями [29]. НKK используется для решения задач, связанных с определением и оценкой влияния факторов ситуации, а также для получения прогнозов развития ситуации на основе вычисленных влияний. Основные элементы, необходимые при создании НKK: узлы, связи, матрица весов связей, функция активации, обучение и обновление, распространение активации.

Другими известными примерами когнитивных карт выступают *Spoofing*, *Tampering*, *Repudiation*, *Information Disclosure*, *Denial of Service*, *Elevation of Privilege (STRIDE)* и *Process for Attack Simulation and Threat Analysis (PASTA)* [30–33].

STRIDE направлен на идентификацию, классификацию и последующую нейтрализацию потенциальных угроз. Рассматривается шесть основных типов угроз, учитываемых в *STRIDE*.

1. *Spoofing* (Подделывание идентичности). Атака, при которой злоумышленники пытаются выдавать себя за доверенного пользователя или систему.
2. *Tampering* (Манипулирование данными). Угроза нарушения целостности данных путём несанкционированного изменения их содержания.
3. *Repudiation* (Отказ от обязательств). Угроза, направленная на уязвимость, позволяющую пользователям отрицать совершение определенных действий.
4. *Information Disclosure* (Раскрытие конфиденциальной информации). Угроза утечки конфиденциальной информации вследствие неправильного конфигурирования систем или наличия уязвимых точек входа.
5. *Denial of Service* (Отказ в обслуживании). Угроза снижения доступности ресурса для легитимных пользователей, приводящая к перегрузке или выходу системы из строя.
6. *Elevation of Privilege* (Повышение привилегий). Угроза получения злоумышленником большего набора прав и возможностей в системе, чем предусмотрено правилами безопасности.

Анализ угроз с использованием *STRIDE* включает несколько ключевых шагов.

1. Идентификация компонентов системы, подлежащих защите.
2. Классификация угроз по одной из шести категорий, приведенных выше.
3. Оценка вероятности возникновения угроз и потенциального ущерба.
4. Выбор мер противодействия каждому типу угроз.
5. Реализация мер защиты.

PASTA позволяет проводить всестороннюю оценку и минимизацию рисков безопасности путем моделирования атак и анализа угроз. Основные этапы процесса *PASTA* следующие.

Этап 1. Сбор и анализ информации. Всестороннее исследование системы с оценкой важнейших активов и потенциальных уязвимых точек.

Этап 2. Имитация атаки. Моделирование реального хода атаки и подбор средств, которые могли бы использовать злоумышленники.

Этап 3. Анализ последствий. Оценка ущерба, который мог бы быть нанесён системой в результате успешной атаки.

Этап 4. Выбор мер защиты. Определение и реализация необходимых защитных мер для минимизации рисков и повышения устойчивости системы.

3. Базы знаний угроз.

Базы знаний угроз содержат информацию о существующих угрозах, уязвимостях и методах атак, делая выявление и устранение рисков более оперативным. Наиболее известными примерами баз знаний угроз являются: *MITRE Adversarial Tactics, Techniques & Common Knowledge (ATT&CK)* и *Common Vulnerability Scoring System (CVSS)* [34, 35].

MITRE ATT&CK служит для описания поведения злоумышленников и их тактик, техник и процедур при проведении кибератак. Она состоит из трех основных матриц, каждая из которых ориентирована на определенный тип целевого окружения.

1. *Enterprise Matrix*. Охватывает техники атак на рабочие станции и серверы, функционирующие под управлением операционных систем *Windows*, *macOS*, *Linux* и других платформ.

2. *Mobile Matrix*. Фокусируется на методах атак на мобильные устройства *Android* и *iOS*.

3. *PRE-ATT&CK Matrix*. Описывает подготовительный этап перед осуществлением атаки, включая разведку угроз и подготовку инфраструктуры.

CVSS позволяет измерить и сравнить уязвимости, присваивая каждой из них единую оценку, варьирующуюся от нуля до десяти, со следующими соответствиями:

- ♦ 0–3,9 – низкий уровень риска;
- ♦ 4,0–6,9 – средний уровень риска;
- ♦ 7,0–8,9 – высокий уровень риска;
- ♦ 9,0–10 – чрезвычайно высокий уровень риска.

Сначала рассчитывается базовая оценка, включающая доступность, воздействие, степень потенциального ущерба атаки и прочие факторы. Затем применяются поправки временной оценки, такие как наличие работающего средства исправлений уязвимости, их доступность и уровень осведомленности о проблеме. Наконец, дополнительно уточняют значение экологическими показателями: важностью системы, критичностью данных, значимостью возможного нарушения функционирования системы.

Несмотря на преимущества описанных моделей, они обладают рядом значимых недостатков. Онтологические модели хоть и основаны на принципах формальной строгости и предопределенные правила, но характеризуются низкой адаптивностью и масштабируемостью. Когнитивные карты угроз хорошо подходят для визуализации и первичного анализа направлений атак, но ограничены в возможностях по формированию сложных логических конструкций и выполнению автоматизированного анализа. Базы знаний угроз, такие как *MITRE ATT&CK* и *CVSS*, собирают обширную информацию о текущих угрозах и уровнях риска, но не предоставляют инструментов для оперативной аналитики и моделирования сценариев угроз.

Использование предлагаемой семантической является, в данном случае, более предпочтительным, поскольку она:

- ♦ расширяет таксономию концептов, необходимых для оценки устойчивости и управления рисками с учетом функциональных зависимостей между ними;
- ♦ обеспечивает полноту учета факторов безопасности, позволяющих получать более детализированную оценку средств защиты при проектировании ИТКС;
- ♦ реализует адаптивное управление рисками, создавая условия для комплексного мониторинга киберустойчивости и бизнес-процессов в ходе эксплуатации ИТКС;
- ♦ включает в себя формализованную целевую функцию системы защиты ИТКС, отражающую основные свойства безопасности, предъявляемые к информационным потокам и функциям бизнес-процессов.

Рассмотрим концептуальную модель (КМ), представленную на рис. 1 и разработанную с использованием аппарата семантических сетей. Первоначально она была представлена в работе [36]. В данном исследовании модель была переосмыслена и модифицирована с точки зрения оценки киберустойчивости.

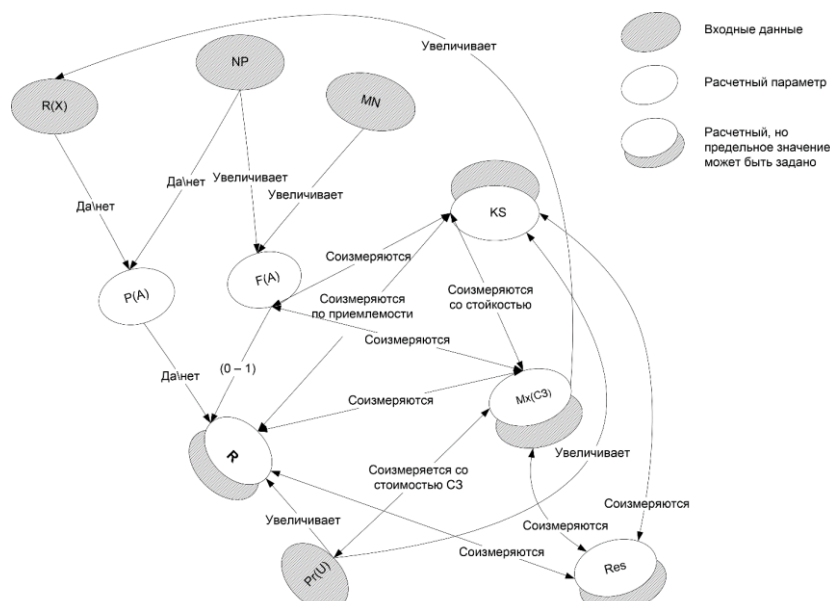


Рис. 1. Концептуальная модель влияния факторов безопасности ИТКС

Здесь вершины графа интерпретируются следующим образом:

1. NP – потенциал нарушителя, оцениваемый на основании его возможностей;
2. $R(X)$ – рейтинг уязвимости компонентов системы, составленный на основе оценок частоты использования уязвимости нарушителем и трудности ее обнаружения;
3. MN – мотивация нарушителя, оцениваемая экспертами;
4. $P(A)$ – возможность осуществления угрозы (атаки) нарушителем, который обладает потенциалом NP и сможет воспользоваться уязвимостью с рейтингом $R(X)$;
5. $F(A)$ – сила (мощность) атаки, ассоциируемая с ресурсами (усилиями), необходимыми для преодоления стойкости защитных механизмов;
6. $\bar{KS} = \{K, C, D\}$ – целевая функция безопасности, которую должна обеспечивать система защиты ИТКС на заданном уровне, представляет собой вектор основных свойств безопасности, предъявляемых к информационным потокам и функциям бизнес-процессов предприятия. Для измерения компонентов вектора $\bar{KS}$ вводятся оценочные шкалы «Уровень конфиденциальности» (K), «Уровень целостности» (C) и «Уровень доступности» (D) в пространстве лингвистических переменных;
7. $Pr(U)$ – ценность информационных ресурсов, используемых при выполнении функций бизнес-процесса. Ценность ресурсов определяет и ущерб, наносимый атакой;
8. R – риск получения ущерба $Pr(U)$ в случае нарушения критериев $\bar{KS}$ и осуществления атаки; риск выступает интегральным фактором-посредником, связывающим угрозы, ценность актива ($Pr(U)$) и требования безопасности.
9. $Mx(C3)$ – защитные механизмы, реализуемые программно-аппаратными средствами;
10. Res – устойчивость информационной системы к нарушению вектора $\bar{KS}$.

Если потенциал нарушителя, оцениваемый его возможностями, превосходит рейтинг уязвимостей $NP > R(X)$, то атака A , в принципе, возможна. Но данные отношения могут принимать значения «да\нет», т.е. возможна\невозможна. Чтобы оценить степень возможности, надо рассмотреть и другие факторы. В частности, потенциал нарушителя, т.е. его степень профессионализма, знания о системе, используемые средства и т.п. напрямую влияют на силу атаки ($<NP, \text{«увеличивает»}, F(A)>$) (здесь и далее влияние измеряется экспертной нормированной шкалой от нуля до единицы). На силу атаки огромное влияние имеет также мотивация нарушителя MN . Если она велика, то, и, не имея собственных возможностей, он может нанять профессионала, поэтому имеет место связь $<MN, \text{«увеличивает»}, F(A)>$.

Влияние концепта $Pr(U)$ (ценность/ущерб) на систему представленных в КК понятий сказывается следующим образом:

а) ценностью определяется требуемый уровень безопасности информационной системы ($\langle Pr(U), \text{«увеличивает»}, \bar{KS} \rangle$);

б) от величины оценки ущерба зависит показатель риска R ($\langle Pr(U), \text{«увеличивает»}, R \rangle$);

в) потенциальный ущерб следует соизмерять со стоимостью системы защиты, по крайней мере, для коммерческих предприятий ($\langle Pr(U), \text{«соизмеряется»}, Mx(C3) \rangle$).

Механизмы защиты Mx , с одной стороны, обеспечивают безопасность и выбирают-ся в соответствии с уровнем критериев ($\langle Mx(C3), \text{«соизмеряется»}, \bar{KS} \rangle$), а с другой – противостоят силе атаки $F(A)$. А это значит, что силу атаки, в силу транзитивности, можно структурировать измерительными шкалами критериев безопасности и имеет место отношение $\langle F(A), \text{«соизмеряется»}, \bar{KS} \rangle$.

Величина рисков предприятия R при нарушении уровня зависит от трех причин:

а) возможности осуществимости атаки $P(A)$ ($\langle P(A), \text{«да/нет»}, R \rangle$);

б) величины оценки ущерба ($\langle Pr(U), \text{«увеличивает»}, R \rangle$);

в) силы атаки ($\langle F(A), \text{«(0 – 1)»}, R \rangle$).

В свою очередь, показатель риска R оказывает влияние на выбор $C3$: чем выше величина риска, тем более стойкие механизмы и $C3$ надо выбирать и наоборот ($\langle R, \text{«соизмеряется»}, Mx(C3) \rangle$).

Управление рисками реализуется двусторонней связью $\langle \bar{KS}, \text{«соизмеряются»}, R \rangle$. Семантика связи заключается в том, чтобы после выбора Mx , обеспечивающих требуемый уровень $\bar{KS}$, произвести оценку будущего риска при данном наборе защитных средств (это можно сделать благодаря связи $\langle Mx(C3), \text{«увеличивает»}, R(X) \rangle$, которая говорит о том, что средства защиты выбранного набора изменяют величину исходного риска) и сделать вывод о его приемлемости. Если риск неприемлем, берется другой набор $Mx(C3)$ и производится сближение оценок и по рискам R , и по уровню безопасности.

Устойчивость Res напрямую зависит от используемых в системе защиты ИТКС защитных механизмов $Mx(C3)$ ($\langle Res, \text{«соизмеряется»}, Mx(C3) \rangle$). При этом требования к уровню устойчивости предъявляются исходя из установленного $\bar{KS}$.

Вывод об уровне устойчивости ИТКС строится на основании оценивания риска (приемлем/неприемлем). Для этого после выбора $Mx(C3)$, обеспечивающих требуемый уровень $\bar{KS}$, необходимо произвести оценку будущего риска при данном наборе защитных средств (это можно сделать благодаря связи $\langle Mx(C3), \text{«соизмеряется»}, R(X) \rangle$, которая говорит о том, что средства защиты выбранного набора изменяют величину исходного риска) и сделать вывод о его приемлемости. Если риск неприемлем, берется другой набор $Mx(C3)$ и производится сближение оценок и по рискам R , и по уровню безопасности $\bar{KS}$.

На КМ можно выделить несколько контуров, которые отражают различные парадигмы управления безопасностью при организации защиты ИТКС. Каждый контур образован двунаправленными отношениями. Это означает, что происходит взаимовлияние факторов и, стало быть, выбор той или иной вершины в качестве цели при принятии решения по выбору вариантов определит направление стрелок и выбор генеральной стратегии политики безопасности предприятия. Ниже описаны некоторые из них:

Контур 1. Образован вершинами $F(A)$, $\bar{KS}$, $Mx(C3)$. Отражает идею управления безопасностью с двух точек зрения: классической (на основании модели нарушителя) и процессной.

Контур 2. Вершины $F(A)$, R , $Mx(C3)$. Отражает политику, требующую разработки повышенных требований к безопасности, основанную на управлении рисками [37].

Контур 3. Вершины $\bar{KS}$, R , $Mx(C3)$. Отражает идею управления безопасностью ИТКС, исходя из оптимального баланса остаточных рисков и уровня безопасности.

Контур 4. Вершины $\bar{KS}$, R , $Mx(C3)$, Res . Отражает комплексный подход к управлению безопасностью ИТКС с учетом требуемого уровня безопасности, а также оценок устойчивости и рисков.

Вершина *Res* наглядно показывает взаимосвязи устойчивости с остальными концептами предметной области (как прямые, так и косвенные). Она представляет собой эмерджентное свойство всей системы, возникающее из взаимодействия перечисленных факторов. Это позволяет выбрать оптимальную политику управления при некотором требуемом уровне устойчивости, что особенно важно при обеспечении безопасности ИТКС предприятия.

Применение метода оценки параметров оконечных устройств. Механизмы защиты от информационных угроз, представленные в КМ концептом $Mx(C3)$, напрямую влияют на устойчивость как ИТКС в целом, так и отдельных ее компонентов. В работе [38] представлена следующая классификация защитных средств и механизмов на группы.

Группа 1: основные (или целевые). Механизмы, обеспечивающие основные свойства безопасности *K*, *C*, *D*. Используют дискреционный метод доступа, мандатный метод доступа, ролевое управление доступом и т.д.

Группа 2: обеспечивающие. Механизмы, осуществляющие дополнительные действия, необходимые для функционирования целевых *Mx* на том или ином уровне безопасности (например, идентификация, аутентификация, криптографическая поддержка, межсетевые экраны и т.д.)

Группа 3: Управляющие. Механизмы, которые обеспечивают согласованное функционирование *Mx* первой и второй групп:

- ◆ мониторинг и аудит событий, происходящих в системе;
- ◆ анализ защищенности, т.е. выявление и анализ уязвимостей;
- ◆ оценка рисков, сопровождающих функционирование системы;
- ◆ администрирование системы безопасности:

В работах [23, 24] авторами был разработан метод оценки корректности параметров оконечных устройств (устройств нижнего уровня) АСУ ТП, являющийся, по сути, компонентом средства мониторинга и анализа защищенности (САЗ) и принадлежащий к группе 3 вышеуказанной классификации (управляющий механизм защиты). Нижний уровень автоматизированных систем управления технологическими процессами, как правило, представлен датчиками, счетчиками, исполнительными механизмами, а также контроллерами, передающими данные от указанных оконечных устройств далее, в верхние уровни системы. Алгоритм предполагает создание комплексных отпечатков (КО) датчиков и их последующее сравнение. Комплексный отпечаток состоит из цифрового отпечатка, основанного на динамической модели сигнала датчика, и аналогового отпечатка, который строится на основе параметров сигнала датчика на технологической шине. Посредством сравнения эталонного КО устройства с текущим КО (с определенной периодичностью или по запросу оператора) можно сделать вывод о подмене или неисправности устройства, т.е. оценить корректность данных, им передаваемых. Предполагается, что при неудовлетворительной оценке осуществляется необходимое управляющее воздействие (как правило, физическая проверка датчика, работы по его ремонту или замене и т.д.).

Рассмотрим использование описанного метода в управленческом контуре 4.

Алгоритм метода, будучи экземпляром $Mx(C3)$ группы 3, осуществляет анализ уровня защищенности и его влияние на уровень киберустойчивости системы (*Res*), поскольку направлен на своевременное обнаружение некорректного сигнала от датчика. Такой сигнал может быть свидетельством как воздействия внутреннего нарушителя, так и выхода устройства из нормального режима работы в результате повреждения, устаревания и т.д.

Использование рассматриваемого алгоритма снижает риски (*R*) возникновения следующих событий:

- ◆ подмена сигнала устройства внутренним нарушителем;
- ◆ подмена самого устройства внутренним нарушителем;
- ◆ выход устройства из строя вследствие неисправности или устаревания.

Величина, на которую снижаются эти риски, определяется исходя из «жесткости» требований, задаваемых $\bar{K}\bar{S}$. Можно настроить систему поддержки принятия решений с данным САЗ так, что она будет требовать проверку датчика при даже самом малом отклонении от эталонного отпечатка: это значительно понизит риски, но при этом может быть экономически нецелесообразно. Поэтому так важно верно определить $\bar{K}\bar{S}$ исходя из необходимого уровня защиты системы.

Значение вектора $\bar{K}\bar{S}$ опосредованно определяет следующее.

1. Диапазоны значений шкалы оценки отклонения от эталонного отпечатка. В зависимости от требуемого уровня безопасности могут быть заданы разные интервалы для значений «норма», «рекомендуется проверка», «требуется проверка». Например, при высоких требованиях $\bar{K}\bar{S}$ в системах, относящихся к критическим информационным инфраструктурам, шкала оценки отклонения может принимать вид, представленный в табл. 1.

Таблица 1

Шкала оценки отклонения от эталонного отпечатка (пример)

Величина отклонения, %	Оценка
0-1	норма
1-5	рекомендуется проверка
5-100	требуется проверка

2. Частота обновления эталонных отпечатков. Чем выше требования по безопасности, тем чаще необходимо актуализировать эталонные отпечатки, чтобы величина отклонения при штатном функционировании датчика не превышала установленное пороговое значение для «нормы» (к примеру, вследствие сезонности).

3. Частота сравнения отпечатков. Здесь высокие требования по безопасности могут означать, что устройства нужно опознавать по их текущим комплексным отпечаткам как можно чаще.

Из рис. 1 видно, что все вышеупомянутые концепты КМ оказывают прямое влияние на работу алгоритма; в зависимости от требований и условий работы его параметры могут быть изменены.

Заключение. Из всего многообразия видов устойчивости систем (структурная-топологическая, организационная, функциональная, процессная и т.п.) для исследования в рамках данной работы была выбрана устойчивость информационной системы к различным видам компрометаций и факторам неблагоприятных условий и инцидентов с точки зрения информационной безопасности.

Результаты, представленные в статье, позволяют осуществлять оценку влияния различных факторов информационной безопасности на уровень риска и киберустойчивости ИТКС. Для этого была создана семантическая концептуальная модель влияния факторов информационной безопасности, таких как рейтинг уязвимостей программного обеспечения, возможности нарушителя и его мотивы, свойства безопасности в составе целевого вектора $\bar{K}\bar{S}$, механизмы и средства защиты, уровень киберустойчивости и т.д. На основе данной модели появляется возможность прогнозировать уровень риска/киберустойчивости при воздействиях различных факторов безопасности систем защиты ИТКС. Для этого выделены несколько контуров управления. Следует отметить, что разработанная модель отражает принципы процессно-ориентированного подхода к защите информационных систем.

Показано, что метод оценки корректности параметров оконечных устройств систем управления технологическими процессами в данной модели реализуется как экземпляр управляющих механизмов концепта модели « $Mx(C3)$ » во множестве средств защиты, для чего приведена классификация механизмов защиты. Кроме того, представлена процедура, демонстрирующая встраивание данного метода в контекст модели.

В перспективе авторы планируют продолжение исследований в области теоретической проработки управленческого аспекта модели, а также ее программной реализации как средства поддержки принятия решений при проектировании и эксплуатации систем обеспечения безопасности ИТКС.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Максимова Е.А., Садовникова Н.П. Оценка инфраструктурной устойчивости субъекта критической информационной инфраструктуры при деструктивных воздействиях // Известия ЮФУ. Технические науки. – 2021. – № 4 (221). – С. 155-165. – DOI: 10.18522/2311-3103-2021-4-155-165.
2. Ross R., Johnson L. Guide for Assessing the Security Controls in Federal Information Systems and Organizations: Building Effective Security Assessment Plans, Special Publication (NIST SP). – Gaithersburg MD: National Institute of Standards and Technology. 2010. – 399 p.
3. Воеводин В.А. Модель оценки функциональной устойчивости элементов информационной инфраструктуры для условий воздействия множества компьютерных атак // Информатика и автоматизация. – СПб.: ФИЦ РАН, 2023. – Т. 22, № 3. – С. 691-715. – DOI: 10.15622/ia.22.3.8.
4. Тарасов А.А. Функциональная устойчивость компьютерных систем как фактор обеспечения их информационной безопасности // Информационная безопасность России в условиях глобального информационного общества: Сб. материалов 3-й Всерос. конф. – М., 2002. – С. 193-200.
5. Бородакий Ю.В., Лободинский Ю.Г. Информационные технологии. Методы, процессы, системы. – М.: Радио и связь, 2002. – 451 с. – ISBN 5-256-01566-4.
6. Королев А.Н. Функциональная устойчивость навигационно-информационных систем // Известия вузов. Приборостроение. – 2018. – Т. 61, № 7. – С. 559-565. – DOI: 10.17586/0021-3454-2018-61-7-559-565.
7. Бородакий Ю.В., Тарасов А.А. О функциональной устойчивости информационно-вычислительных систем // Известия ЮФУ. Технические науки. – 2006. – № 7. – С. 5-12.
8. Лукинова О.В. Компьютерные методы и алгоритмы управления безопасностью информационных систем. – М.: ИПУ РАН, 2014. – 248 с.
9. Лукинова О.В. Формализация задачи планирования защиты распределенной компьютерной сети на основе бизнес-процессного подхода // Надежность. – 2009. – № 1. – С. 72-80.
10. Лукинова О.В. Формирование модели угроз безопасности компьютерной сети при бизнес-процессном подходе // Реинжиниринг бизнес-процессов на основе современных информационных технологий: Тр. XII науч.-практ. конф. – М., 2009. – С. 170-176.
11. ГОСТ Р 51897–2011/Руководство ИСО 73:2009. Менеджмент риска. Термины и определения. – М.: Стандартиформ, 2012. – 16 с.
12. ГОСТ Р ИСО/МЭК 13335-1–2006. Информационная технология. Методы и средства обеспечения безопасности. Ч. 1. Концепция и модели менеджмента безопасности информационных и телекоммуникационных технологий. – М.: Стандартиформ, 2007. – 23 с.
13. Азбука киберустойчивости. – 2025. – URL: <https://www.kaspersky.ru/blog/cyber-resilience-101/39564/> (дата обращения: 03.11.2025).
14. Краснов А.Е., Мосолов А.С., Феоктистова Н.А. Оценивание устойчивости критических информационных инфраструктур к угрозам информационной безопасности // Безопасность информационных технологий. – 2021. – Т. 28, № 1. – С. 106-120. – DOI: 10.26583/bit.2021.1.09.
15. Солянов Д.А., Тимирязова Д.Р. Стратегии повышения уровня киберустойчивости в корпоративной инфраструктуре // Международный студенческий научный вестник. – 2025. – № 1. – URL: <https://eduherald.ru/ru/article/view?id=21744> (дата обращения: 03.11.2025).
16. NIST SP 800-39. Managing Information Security Risk: Organization, Mission, and Information System View / National Institute of Standards and Technology. – Gaithersburg: NIST, 2011. – 82 p. – URL: <https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-39.pdf> (дата обращения: 03.11.2025).
17. Российская Федерация. Законы. О персональных данных : Федер. закон № 152-ФЗ: принят Гос. Думой 8 июля 2006 г.: одобр. Советом Федерации 14 июля 2006 г. – М., 2006. – Доступ из справ.-правовой системы «КонсультантПлюс».
18. Российская Федерация. Законы. О безопасности : Федер. закон № 390-ФЗ: принят Гос. Думой 7 декабря 2010 г. : одобр. Советом Федерации 15 декабря 2010 г. – М., [2010. – Доступ из справ.-правовой системы «КонсультантПлюс».
19. Российская Федерация. Законы. Об информации, информационных технологиях и о защите информации: Федер. закон № 149-ФЗ: принят Гос. Думой 8 июля 2006 г.: одобр. Советом Федерации 14 июля 2006 г. – М., 2006. – Доступ из справ.-правовой системы «КонсультантПлюс».
20. ITIL Essentials for IT Service Management: Матер. учебного курса. – Версия В.00. – Hewlett-Packard Education, 1998. – 178 с.
21. ГОСТ Р ИСО/МЭК 31010-2011. Менеджмент риска. Методы оценки риска. – Введ. 2012–12–01. – М.: Стандартиформ, 2012. – 74 с.
22. ISO/IEC 27005:2018. Information technology – Security techniques – Information security risk management. – 3rd ed. – Geneva: ISO/IEC, 2018. – 61 p.

23. Богачева Д.Н., Лукинова О.В. Вопросы оценки корректности данных устройств нижнего уровня автоматизированных систем // Матер. 32-й Международной научно-технической конференции «Системы безопасности – 2023». – М.: Академия ГПС МЧС России, 2023. – С. 349-355.
24. Bogacheva D.N., Lukinova O.V., Pavlova E.S. The Approach to Assessing the Correctness of Automated System Endpoint Devices' Parameters Using their Reference Models // Proceedings of the 2024 International Russian Smart Industry Conference (SmartIndustryCon). – Sochi: IEEE, 2024. – P. 850-854.
25. Массель А.Г., Гаськова Д.А. Онтологический инжиниринг для разработки интеллектуальной системы анализа угроз и оценки рисков кибербезопасности энергетических объектов // Онтология проектирования. – 2019. – Т. 9, № 2 (32). – С. 225-238.
26. Бурый А.С., Усцелемов В.Н. Онтологический подход к формированию когнитивных моделей оценки кибербезопасности // Информационно-экономические аспекты стандартизации и технического регулирования. – 2020. – № 3 (55). – С. 77-84.
27. Ученые МГУ разработали новую онтологию информационной безопасности. – 2025. – URL: <https://cs.msu.ru/news/3933> (дата обращения: 03.11.2025).
28. Колесникова Д.С., Вережагина Е.А. Применение онтологий в обучающих системах // Инженерный вестник Дона. – № 6 (102). – С. 247-257.
29. Гузайров М.Б., Вульфин А.М., Картак В.М., Кириллова А.Д., Миронов К.В. Сравнительный анализ алгоритмов когнитивного моделирования при оценке рисков информационной безопасности // Тр. ИСА РАН. – 2019. – Т. 69, № 4. – URL: <http://www.isa.ru/proceedings/images/documents/2019-69-4/62-69.pdf> (дата обращения: 03.11.2025).
30. Яцук К.В., Свиридов О.И., Иванов Д.А., Скоробогатов С.Ю. Методика моделирования угроз STRIDE на технологию SDN-контроллеров // Известия Тульского государственного университета. Технические науки. – 2022. – № 3. – С. 347-352.
31. Khan R., McLaughlin K., Lavery D., Sezer S. STRIDE-based Threat Modeling for Cyber-Physical Systems // Proceedings of the 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe. – 2017. – P. 1-6.
32. Shevchenko N., Chick T.A., O'Riordan P., Scanlon T.P., Woody C. Threat modelling: A summary of available methods. – Carnegie Mellon University Software Engineering Institute. – 2018. – С. 1-24.
33. Миняев А.А. Моделирование угроз безопасности информации в территориально-распределенных информационных системах // Научные технологии в космических исследованиях Земли. – 2021. – № 13 (2). – С. 52-65.
34. Зуфарова А.С., Кошелева А.Д. Что такое MITRE ATT&CK: разбор популярной тактической модели. Информатика. Экономика. Управление // Informatics. Economics. Management. – 2025. – № 4 (1). – С. 2027-2037. – DOI: 10.47813/2782-5280-2025-4-1-2027-2037.
35. Дойникова Е.В., Чечулин А.А., Котенко И.В. Оценка защищенности компьютерных сетей на основе метрик CVSS // Информационно-управляющие системы. – 2017. – №. 6 (91). – С. 76-87.
36. Бойченко А.В., Лукинова О.В. Когнитивный подход к анализу влияния факторов информационной безопасности // Тр. 7-й Международной научно-практической конференции «Интегрированные модели и мягкие вычисления в искусственном интеллекте». – М.: Физматлит, 2013. – С. 583-586.
37. Афонцев Э.В., Поршнев С.В. Опыт построения методик обнаружения вирусной сетевой активности // Вестник УГТУ-УПИ. – 2004. – № 20 (50). – С. 215-217.
38. Лукинова О.В., Пугачев А.В. Особенности построения профилей систем безопасности ИС // Открытое образование. – 2015. – № 4. – С. 80-87. – DOI: 10.21686/1818-4243-2015-4(111-38-44).

REFERENCES

1. Maksimova E.A., Sadovnikova N.P. Otsenka infrastrukturnoy ustoychivosti sub'yekta kriticheskoy informatsionnoy infrastruktury pri destruktivnykh vozdeystviyakh [Assessment of the infrastructure resilience of a critical information infrastructure subject under destructive impacts], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 4, pp. 155-165, DOI: 10.18522/2311-3103-2021-4-155-165.
2. Ross R., Johnson L. Guide for Assessing the Security Controls in Federal Information Systems and Organizations: Building Effective Security Assessment Plans, Special Publication (NIST SP), Gaithersburg MD, National Institute of Standards and Technology, 2010, 399 p.
3. Voevodin V.A. Model' otsenki funktsional'noy ustoychivosti elementov informatsionnoy infrastruktury dlya usloviy vozdeystviya mnozhestva komp'yuternykh atak [Model for assessing the functional stability of information infrastructure elements under multiple computer attacks], *Informatika i avtomatizatsiya* [Informatics and Automation], Saint Petersburg, FITs RAN, 2023, Vol. 22, No. 3, pp. 691-715. DOI: 10.15622/ia.22.3.8.

4. Tarasov A.A. Funktsional'naya ustoychivost' komp'yuternykh sistem kak faktor obespecheniya ikh informatsionnoy bezopasnosti [Functional stability of computer systems as a factor in ensuring their information security], *Informatsionnaya bezopasnost' Rossii v usloviyakh global'nogo informatsionnogo obshchestva* [Information Security of Russia in the Global Information Society], Moscow, 2002, pp. 193-200.
5. Borodakiy Yu.V., Lobodinskiy Yu.G. Informatsionnye tekhnologii. Metody, protsessy, sistemy [Information technologies. Methods, processes, systems], Moscow, Radio i svyaz', 2002, 451 p.
6. Korolev A.N. Funktsional'naya ustoychivost' navigatsionno-informatsionnykh sistem [Functional stability of navigation and information systems], *Izvestiya vuzov. Priborostroenie* [Journal of Instrument Engineering], 2018, Vol. 61, No. 7, P. 559–565. DOI: 10.17586/0021-3454-2018-61-7-559-565.
7. Borodakiy Yu.V., Tarasov A.A. O funktsional'noy ustoychivosti informatsionno-vychislitel'nykh sistem [On the functional stability of information and computing systems], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2006, No. 7, pp. 5-12.
8. Lukinova O.V. Komp'yuternye metody i algoritmy upravleniya bezopasnost'yu informatsionnykh sistem [Computer methods and algorithms for information systems security management], Moscow, IPU RAN [ICS RAS], 2014, 248 p.
9. Lukinova O.V. Formalizatsiya zadachi planirovaniya zashchity raspredelennoy komp'yuternoy seti na osnove biznes-protsessnogo podkhoda [Formalization of the task of planning the protection of a distributed computer network based on a business process approach], *Nadezhnost'* [Dependability], 2009, No. 1, pp. 72-80.
10. Lukinova O.V. Formirovanie modeli ugroz bezopasnosti komp'yuternoy seti pri biznes-protsessnom podkhode [Formation of a computer network security threat model using a business process approach], *Reinzhiniring biznes-protsessov na osnove sovremennykh informatsionnykh tekhnologiy* [Business Process Reengineering Based on Modern Information Technologies], Moscow, 2009, pp. 170-176.
11. GOST R 51897-2011/ISO Guide 73:2009. Menedzhment riska. Terminy i opredeleniya [Risk management. Vocabulary. Guidelines for use in standards]. Moscow: Standartinform, 2012, 16 p.
12. GOST R ISO/IEC 13335-1-2006. Informatsionnaya tekhnologiya. Metody i sredstva obespecheniya bezopasnosti. Chast' 1. Kontseptsiya i modeli menedzhmenta bezopasnosti informatsionnykh i telekommunikatsionnykh tekhnologiy [Information technology. Security techniques. Part 1. Concepts and models for information and telecommunications technology security management], Moscow, Standartinform, 2007, 23 p.
13. Azbuka kiberustoychivosti [Cyber resilience ABCs], *Kaspersky Blog* [Kaspersky Blog], 2025. Available at: <https://www.kaspersky.ru/blog/cyber-resilience-101/39564/> (accessed 03 November 2025).
14. Krasnov A.E., Mosolov A.S., Feoktistova N.A. Otsenivanie ustoychivosti kriticheskikh informatsionnykh infrastruktur k ugrozam informatsionnoy bezopasnosti [Evaluating the resilience of critical information infrastructures to information security threats], *Bezopasnost' informatsionnykh tekhnologiy* [IT Security], 2021, Vol. 28, No. 1, pp. 106-120. DOI: 10.26583/bit.2021.1.09.
15. Solyanov D.A., Timiryanova D.R. Strategii povysheniya urovnya kiberustoychivosti v korporativnoy infrastrukture [Strategies for increasing the level of cyber resilience in corporate infrastructure], *Mezhdunarodnyy studencheskiy nauchnyy vestnik* [International Student Scientific Bulletin], 2025, No. 1. Available at: <https://eduherald.ru/ru/article/view?id=21744> (accessed 03 November 2025).
16. NIST SP 800-39. Managing Information Security Risk: Organization, Mission, and Information System View, *National Institute of Standards and Technology*, Gaithersburg, NIST, 2011, 82 p. Available at: <https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-39.pdf> (accessed 03 November 2025).
17. Russian Federation. Laws. O personal'nykh dannykh: Feder. zakon No. 152-FZ [On Personal Data: Federal Law No. 152-FZ]. Moscow, 2006, ConsultantPlus.
18. Russian Federation. Laws. O bezopasnosti: Feder. zakon No. 390-FZ [On Security: Federal Law No. 390-FZ]. Moscow, 2010, ConsultantPlus.
19. Russian Federation. Laws. Ob informatsii, informatsionnykh tekhnologiyakh i o zashchite informatsii: Feder. zakon No. 149-FZ [On Information, Information Technologies and Information Protection: Federal Law No. 149-FZ]. Moscow, 2006, ConsultantPlus.
20. ITIL Essentials for IT Service Management: materialy uchebnogo kursa. Versiya B.00 [ITIL Essentials for IT Service Management: Course Materials. Version B.00], *Hewlett-Packard Education*, 1998, 178 p.
21. GOST R ISO/IEC 31010-2011. Menedzhment riska. Metody otsenki riska [Risk management. Risk assessment techniques]. Moscow, Standartinform, 2012, 74 p.
22. ISO/IEC 27005:2018. Information technology — Security techniques — Information security risk management, 3rd ed., Geneva, ISO/IEC, 2018, 61 p.
23. Bogacheva D.N., Lukinova O.V. Voprosy otsenki korrektnosti dannykh ustroystv nizhnego urovnya avtomatizirovannykh sistem [Issues of assessing the correctness of data from low-level devices of automated systems], *Materialy 32-y Mezhdunarodnoy nauchno-tekhnicheskoy konferentsii "Sistemy bezopasnosti – 2023"* [Proceedings of the 32nd International Scientific and Technical Conference "Security Systems – 2023"]. Moscow, Akademiya GPS MCHS Rossii, 2023, pp. 349-355.

24. Bogacheva D.N., Lukinova O.V., Pavlova E.S. The Approach to Assessing the Correctness of Automated System Endpoint Devices' Parameters Using their Reference Models, *Proceedings of the 2024 International Russian Smart Industry Conference (SmartIndustryCon)*, Sochi, IEEE, 2024, pp. 850-854.
25. Massel' A.G., Gaskova D.A. Ontologicheskii inzhiniring dlya razrabotki intellektual'noy sistemy analiza ugroz i otsenki riskov kiberbezopasnosti energeticheskikh ob'yektov [Ontological engineering for the development of an intelligent system for threat analysis and cybersecurity risk assessment of energy facilities], *Ontologiya proektirovaniya* [Ontology of Designing], 2019, Vol. 9, No. 2 (32), pp. 225-238.
26. Buryy A.S., Ustselemov V.N. Ontologicheskii podkhod k formirovaniyu kognitivnykh modeley otsenki kiberbezopasnosti [Ontological approach to the formation of cognitive models for cybersecurity assessment], *Informatsionno-ekonomicheskie aspekty standartizatsii i tekhnicheskogo regulirovaniya* [Information and Economic Aspects of Standardization and Technical Regulation], 2020, No. 3 (55), pp. 77-84.
27. Uchenye MGU razrabotali novuyu ontologiyu informatsionnoy bezopasnosti [MSU scientists have developed a new information security ontology], *VMK MSU* [VMK MSU website], 2025. Available at: <https://cs.msu.ru/news/3933> (accessed 03 November 2025).
28. Kolesnikova D.S., Vereshchagina E.A. Primenenie ontologiy v obuchayushchikh sistemakh [Application of ontologies in educational systems], *Inzhenernyy vestnik Dona* [Engineering Journal of Don], 2025, No. 6 (102), pp. 247-257.
29. Guzairov M.B., Vul'fin A.M., Kartak V.M., Kirillova A.D., Mironov K.V. Sravnitel'nyy analiz algoritmov kognitivnogo modelirovaniya pri otsenke riskov informatsionnoy bezopasnosti [Comparative analysis of cognitive modeling algorithms in information security risk assessment], *Trudy ISA RAN* [Proceedings of ISA RAS], 2019, Vol. 69, No. 4. Available at: <http://www.isa.ru/proceedings/images/documents/2019-69-4/62-69.pdf> (accessed 03 November 2025).
30. Yatsuk K.V., Sviridov O.I., Ivanov D.A., Skorobogatov S.Yu. Metodika modelirovaniya ugroz STRIDE na tekhnologiyu SDN-kontrollerov [STRIDE threat modeling methodology for SDN controller technology], *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki* [News of the Tula State University. Technical Sciences], 2022, No. 3, pp. 347-352.
31. Khan R., McLaughlin K., Lavery D., Sezer S. STRIDE-based Threat Modeling for Cyber-Physical Systems, *Proceedings of the 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe*, 2017, pp. 1-6.
32. Shevchenko N., Chick T.A., O'Riordan P., Scanlon T.P., Woody C. Threat modelling: A summary of available methods, *Carnegie Mellon University Software Engineering Institute*, 2018, pp. 1-24.
33. Minyaev A.A. Modelirovanie ugroz bezopasnosti informatsii v territorial'no-raspredelennykh informatsionnykh sistemakh [Information security threat modeling in geographically distributed information systems], *Naukoyemkie tekhnologii v kosmicheskikh issledovaniyakh Zemli* [High-tech Technologies in Earth Space Research], 2021, No. 13 (2), pp. 52-65.
34. Zufarova A.S., Kosheleva A.D. Chto takoe MITRE ATT&CK: razbor populyarnoy takticheskoy modeli [What is MITRE ATT&CK: Analysis of a Popular Tactical Model], *Informatika. Ekonomika. Upravlenie* [Informatics. Economics. Management], 2025, No. 4 (1), pp. 2027-2037. DOI: 10.47813/2782-5280-2025-4-1-2027-2037.
35. Doynikova E.V., Chechulin A.A., Kotenko I.V. Otsenka zashchishchennosti komp'yuternykh setey na osnove metrik CVSS [Security assessment of computer networks based on CVSS metrics], *Informatsionno-upravlyayushchie sistemy* [Information Management Systems], 2017, No. 6 (91), pp. 76-87.
36. Boychenko A.V., Lukinova O.V. Kognitivnyy podkhod k analizu vliyaniya faktorov informatsionnoy bezopasnosti [Cognitive approach to the analysis of the influence of information security factors], *Trudy 7-y Mezhdunarodnoy nauchno-prakticheskoy konferentsii "Integrirovannye modeli i myagkie vychisleniya v iskusstvennom intellekte"* [Proceedings of the 7th International Scientific and Practical Conference "Integrated Models and Soft Computing in Artificial Intelligence"], Moscow, Fizmatlit, 2013, pp. 583-586.
37. Afontsev E.V., Porshnev S.V. Opyt postroeniya metodik obnaruzheniya virusnoy setevoy aktivnosti [Experience in building methods for detecting viral network activity], *Vestnik UGTU-UPI* [Bulletin of USTU-UPI], 2004, No. 20 (50), pp. 215-217.
38. Lukinova O.V., Pugachev A.V. Osobennosti postroeniya profiley sistem bezopasnosti IS [Features of constructing security system profiles for information systems], *Otkrytoe obrazovanie* [Open Education], 2015, No. 4, P. 80-87. DOI: 10.21686/1818-4243-2015-4(111-38-44).

Богачева Дарья Николаевна – Институт проблем управления им. В.А. Трапезникова РАН; e-mail: bogacheva@ipu.ru; г. Москва, Россия; тел.: 84953348910; лаборатория Инфраструктурных систем; м.н.с.

Лукинова Ольга Васильевна – Институт проблем управления им. В.А. Трапезникова РАН; e-mail: lukinova@ipu.ru; г. Москва, Россия; тел.: 84953348910; лаборатория Инфраструктурных систем; д.т.н.; доцент; в.н.с.

Саломатин Александр Александрович – Институт проблем управления им. В.А. Трапезникова РАН; e-mail: sandr@ipu.ru; г. Москва, Россия, тел.: 84953348910; лаборатория Киберфизических систем; к.т.н.; с.н.с.

Bogacheva Darya Nikolaevna – V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences; e-mail: bogacheva@ipu.ru; Moscow, Russia; phone: +74953348910; Infrastructure Systems Laboratory; junior researcher.

Lukinova Olga Vasilievna – V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences; e-mail: lukinova@ipu.ru; Moscow, Russia; phone: +74953348910; Infrastructure Systems Laboratory; dr. of eng. sc.; associate professor; leading researcher.

Salomatin Aleksandr Aleksandrovich – V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences; e-mail: sandr@ipu.ru; Moscow, Russia; phone: +74953348910; Cyber-Physical Systems Laboratory; cand. of eng. sc.; senior researcher.

УДК 004.056.53:004.75

DOI 10.18522/2311-3103-2026-1-179-191

Д.О. Ларин, Р.И. Захарченко, С.А. Диченко

МАНДАТНАЯ АТРИБУТИВНО-РОЛЕВАЯ МОДЕЛЬ УПРАВЛЕНИЯ ДОСТУПОМ В КРУПНОМАСШТАБНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

В условиях стремительного развития информационных систем федерального масштаба, их эволюции в цифровые экосистемы, к процессу обеспечения безопасности обрабатываемой в них информации предъявляются новые требования. Такими требованиями, в частности, являются повышение доступности информации при управлении доступом пользователей с сохранением требуемого уровня ее конфиденциальности, принятие решения о доступе к ресурсам на основе множества факторов. Для их удовлетворения ранее было предложено множество композиционных моделей управления доступом на основе ролей и атрибутов, которые решили ряд актуальных проблем, сохранив удобство администрирования и обеспечив при этом гибкость и масштабируемость без «взрыва ролей». Однако известные модели все еще имеют существенный недостаток – невозможность их использования в информационных системах, где обрабатывается информация высокого уровня значимости. Целью исследования является разработка в рамках методологии субъект-сущностного подхода теории информационной безопасности новой мандатной атрибутивно-ролевой модели управления доступом, а также ее формальное описание с помощью математического аппарата теории автоматов. Использование модели позволит предотвращать загроможденные информационные потоки от объектов с высоким уровнем конфиденциальности к объектам с низким уровнем конфиденциальности динамически, в процессе ограничения набора разрешений, назначенных роли, с помощью реализации мандатного разграничения доступа через отдельную атрибутивную политику, сохраняя при этом возможность предоставления пользователям доступа высокой степени детализации на основе атрибутов контекста. Применение модели может быть востребовано в крупномасштабных информационных системах, где одновременно обрабатывается информация различных уровней конфиденциальности и, ввиду особенностей функционирования, необходима реализация атрибутивного управления доступом.

Атрибутивное управление доступом; ролевое управление доступом, мандатное управление доступом; модель управления доступом; MAP модель; конфиденциальность; доступность.

D.O. Larin, R.I. Zaharchenko, S.A. Dichenko

MANDATORY ROLE-CENTRIC ATTRIBUTE-BASED ACCESS CONTROL MODEL FOR LARGE-SCALE INFORMATION SYSTEMS

In the context of the rapid development of national-scale information systems and their evolution into digital ecosystems, new requirements are imposed on the process of ensuring the security of the information processed within them. These requirements include enhancing information availability in user access management while maintaining the required level of confidentiality, and making access decisions to resources based on multiple factors. To meet these requirements, numerous compositional access control models based on roles and attributes have been proposed previously, which have resolved several pressing issues while maintaining administrative convenience and providing flexibility and scalability

without role explosion. However, known models still have a significant limitation – the impossibility of their use in information systems where high-sensitivity data is processed. The aim of the study is to develop, within the framework of the subject-object approach methodology in information security theory, a new mandatory role-centric attribute-based access control (MRABAC) model, as well as its formal description using the mathematical apparatus of automata theory. The use of the model will enable dynamic prevention of unauthorized information flows from high-confidentiality objects to low-confidentiality objects during the restriction of the permission set assigned to a role, through the implementation of mandatory access control via a separate attribute-based policy, while preserving the ability to provide users with fine-grained access based on contextual attributes. The application of the model may be particularly useful in large-scale information systems where information of various confidentiality levels is processed simultaneously, and, due to operational characteristics, attribute-based access control is necessary.

Attribute-based access control; role-based access control; mandatory access control; access control model; MRABAC; confidentiality; availability.

Введение. Сегодня основной деятельностью как крупных частных компаний, так и целых отраслей государства являются крупномасштабные информационные системы (КМИС), появление которых стало возможным благодаря развитию технологий распределенных вычислений, облачных вычислений, мобильного интернета, интернета вещей. Стремительная эволюция КМИС и их особенности, такие, как большой территориальный размах; многоуровневая иерархическая структура; многоцелевой характер функционирования; разнородность и сложное взаимодействие элементов (информационных ресурсов и систем); огромное количество пользователей, динамический характер их поведения и информационных потребностей [1] порождают необходимость в разработке новых и совершенствовании существующих способов обеспечения безопасности информации и, в частности, моделей и методов управления доступом.

В отличие от информационных систем (ИС) объектового масштаба (одна организация или предприятие), где количество субъектов и объектов доступа ограничено, а требования к управлению доступом статичны и могут быть легко удовлетворены такими классическими моделями управления доступом, как дискреционного (DAC) [2], мандатного (MAC) [3] и ролевого (RBAC) [4] управления доступом или их композицией [5, 6], в современных гетерогенных КМИС доступ зависит от множества факторов, так называемого контекста (context-aware access), в котором поступают запросы на доступ – различных атрибутов субъектов (специализация, возраст, гражданство и др.), объектов (тип, статус, принадлежность владельцу и др.) и среды (местоположение пользователя, время его обращения и др.), и обеспечение такого гибкого доступа выходит за рамки возможностей этих моделей.

Исходя из новых требований к обеспечению доступа пользователей к ресурсам была разработана и принята в качестве стандарта модель атрибутивного управления доступом (ABAC/XACML) [7, 8], которая успешно решает известную проблему «взрыва ролей» модели RBAC и обеспечивает динамическое, детализированное управление доступом (fine-grained access control), но процесс принятия решения о доступе в ней является более сложным и влечет за собой другую известную проблему – проблему «взрыва правил», что делает анализ безопасности ИС трудным, а иногда невозможным.

С целью нивелирования недостатков и эффективного использования взаимодополняющих преимуществ моделей RBAC и ABAC Р. Куном и др. был предложен и получил наибольшее признание в научном сообществе подход к их объединению – атрибутивная модель на основе ролей [9], которая сохраняет преимущества ролевой модели в части удобства администрирования (после определения ролей и разрешений) и упрощения процесса формального анализа безопасности, предотвращая при этом «взрыв ролей» и «взрыв правил» при необходимости учета всех возможных непредвиденных обстоятельств, которые могут возникнуть в процессе обеспечения доступа пользователей к ресурсам.

Однако открытой научно-технической проблемой этой гибридной модели, равно как и моделей RBAC и ABAC по отдельности, остается невозможность их использования в «чистом» виде в КМИС, где одновременно обрабатывается информация различных уровней конфиденциальности и видов (коммерческая, служебная, государственная тай-

на), а пользователи, соответственно, имеют различные права доступа к такой информации. В таких системах предъявляются повышенные требования к обеспечению конфиденциальности информации и контролю информационных потоков [10, 11] и соблюдения их возможно только с помощью реализации дополнительно мандатного управления доступом, которое само по себе является полноценным, отдельным механизмом со своей формализацией.

В опубликованных ранее работах не было представлено формальной модели управления доступом, которая объединяла бы в себе модели RBAC, ABAC и MAC, обеспечивая выполнение трёх известных свойств безопасности по Беллу-ЛаПадуле [3] и сохраняя возможность предоставления доступа высокой степени детализации на основе атрибутов контекста. В работе [12] С. Цзинь и др. представили формальную модель ABAC_ω, способную выразить модели DAC, MAC и RBAC, однако речь в ней идет только об описании конфигурации модели MAC для ее реализации на базе архитектуры XACML. По сути, это все та же классическая модель мандатного управления доступом со своими ограничениями, которую авторы выразили в терминах ABAC. В статье [13] Л. Керр и Д. Алвес-Фосс представили комбинированную модель, объединяющую MAC и ABAC, для систем, где набор атрибутов и возможных значений для них заранее известен и фиксирован. Следовательно, число возможных комбинаций атрибутов относительно невелико, что делает задачу проверки безопасности алгоритмически разрешимой, однако при возникновении необходимости предоставления более детализированного доступа на основе новых атрибутов проблема «взрыва правил» остается актуальной, а наличие в одной политике разрешающих правил как мандатного, так и атрибутивного разграничения доступа, их произвольное комбинирование, делает невозможным строгое формальное обоснование безопасности политик. Также существует ряд известных работ [14–18], в которых авторы развили упомянутый ранее подход Р. Куна и др. и представили различные варианты атрибутивных моделей управления доступом на основе ролей, сосредоточившись на совершенствовании механизмов обеспечения доступности, но в них не затрагиваются вопросы предотвращения запрещенных информационных потоков при обработке в системе информации различных уровней конфиденциальности.

В связи с вышеизложенным особую актуальность приобретает задача логического управления доступом в КМИС с повышенными требованиями как к конфиденциальности, так и к доступности информации путем объединения нескольких разнородных моделей безопасности, каждая из которых имеет формальную модель, задающую строгую семантику и обеспечиваемые свойства безопасности, в единую композиционную модель. «Механическое» объединение нескольких моделей без учета специфики каждой из них и особенностей среды, в которой требуется обеспечить управление доступом, может привести к нарушению их ключевых свойств безопасности и возникновению запрещенных потоков информации в защищаемой системе [19].

Целью данной статьи является представление результатов разработки на основе методологии субъект-сущностного подхода теории информационной безопасности, а также методологии теории автоматов новой мандатной атрибутивно-ролевой модели управления доступом (MAP модель; MRABAC), использование которой позволит сохранить высокий уровень доступности путем повышения степени автоматизации и сокращения времени администрирования системы управления доступом и одновременно обеспечить конфиденциальность информации при необходимости предоставления детализированного доступа на основе атрибутов контекста путем сохранения свойств безопасности моделей MAC и RBAC.

Структура статьи выглядит следующим образом: в разделе 1 дано общее описание разработанной модели, в разделе 2 она описана формально, в разделе 3 представлен пример реализации MAP модели. В Заключение подводятся итоги исследования и освещаются некоторые возможные направления будущей работы по исследуемой проблематике.

1. MAP модель управления доступом. Логической основой для формирования MAP модели (рис. 1) служит атрибутивная модель управления доступом на основе ролей RABAC, представленная С. Цзинем и др. в 2012 году [14], где авторы добавляют филь-

рацию разрешений на операции с объектами на основе атрибутов с целью ограничения набора разрешений роли, доступных для каждой сессии пользователя, не изменяя базовую структуру модели RBAC. В процессе назначения прав активируется сессия s аутентифицированного пользователя u , где пользователю u назначается множество ролей R , на котором задается иерархия в виде отношения частичного порядка « $\leq$ », где вышестоящие роли наследуют разрешения нижестоящих ролей, множеству ролей R назначается множество разрешений P на операции OPS с объектами O . Далее доступное множество разрешений P ограничивается при помощи набора атрибутивных политик фильтрации разрешений PFP , содержащих наборы функций фильтрации $F_1, F_2, \dots, F_n$ на основе конечного набора атрибутов пользователя $UATT$, конечного набора атрибутов объекта $OATT$, после чего вычисляется окончательное доступное множество разрешений P' . Внедрение PFP решает проблему «взрыва ролей» RBAC, сохраняя статические отношения между ролями и разрешениями.

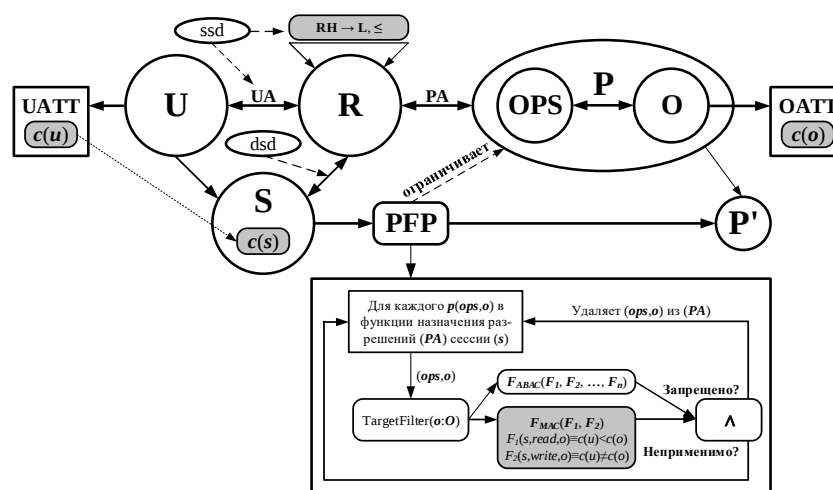


Рис. 1. Структура элементов MAP модели управления доступом

Сокращение времени администрирования системы управления доступом при ее функционировании на базе модели RBAC обеспечивается тем, что комбинация ролевых и атрибутивных политик устраняет необходимость в назначении дополнительных ролей пользователю. Атрибутивные политики покрывают множество условий, одно правило может охватывать множество обновлений ролевой модели и, следовательно, может значительно сократить административные усилия по разработке и обновлению набора ролей, которые являются главным фактором, влияющим на временные задержки при предоставлении доступа. Также, атрибутивный подход обеспечивает возможность доступа любой степени детализации с учетом меняющегося контекста.

Новым в разработанной нами модели является следующее:

- ♦ в набор атрибутов пользователя $UATT$ добавляется атрибут уровня доступа пользователя $c(u)$, в набор атрибутов объекта $OATT$ добавляется атрибут уровня конфиденциальности объекта $c(o)$, активируемой сессии s присваивается уровень доступа $c(s)$, соответствующий уровню доступа пользователя $c(u)$;

- ♦ в иерархии на множестве ролей R реализуются условия, по которым разрешения на операцию чтения наследуются, если уровень конфиденциальности объекта доступа $c(o_1)$ нижестоящей роли не превышает уровень конфиденциальности объекта доступа $c(o_2)$ вышестоящей роли, разрешения на операцию записи от нижестоящей роли к вышестоящей не наследуются, для этой цели в набор атрибутивных политик фильтрации разрешений PFP добавляется набор функций фильтрации, образующих политику F_{MAC} , представляющих собой логические выражения на основе атрибутов $c(u)$, $c(o)$ в виде запрещающих правил на операцию чтения пользователем u объекта o , принадлежащего

множеству объектов O , если уровень конфиденциальности объекта $c(o)$ больше уровня доступа пользователя $c(u)$, запрещающих правил на операцию записи пользователем u в объект o , принадлежащий множеству объектов O , если уровень конфиденциальности объекта $c(o)$ не равен уровню доступа пользователя $c(u)$.

Таким образом реализуется решетка уровней конфиденциальности L в защищаемой системе, из множества разрешений P динамически исключаются разрешения, нарушающие принципы мандатного разграничения доступа, обеспечивается конфиденциальность информации и предотвращаются информационные потоки от объектов с высоким уровнем конфиденциальности к объектам с низким уровнем конфиденциальности.

Стоит отметить, что в MAP модели мы используем строгое мандатное управление доступом, как наиболее безопасный вариант модели Белла-ЛаПадулы.

Определение 1. Доступ пользователя к объекту в рамках активируемой сессии доступа является безопасным для строгого мандатного управления доступом, когда выполняется одно из условий:

- ♦ разрешена операция чтения и $c(u) \geq c(o)$ (ss-свойство);
- ♦ разрешена операция записи в объект o и $c(u) = c(o)$, и, если разрешена операция чтения объекта o' , то $c(o) = c(o')$ (строгое *-свойство) [18].

Обеспечение конфиденциальности информации и предотвращение запрещенных информационных потоков в защищаемой системе при реализации MAP модели осуществляется путем последовательного выполнения подсистемой авторизации системы управления доступом, архитектура которой определяется стандартом XACML, операций по обработке поступающих запросов пользователей на доступ к объектам (ресурсам) и вычислению итогового решения о доступе на основе набора атрибутивных политик фильтрации разрешений PFP , содержащего набор функций фильтрации F_{MAC} строго в виде запрещающих правил, результаты которых имеют приоритет над разрешениями, назначенными ролям.

Приоритет запрещающих решений набора функций фильтрации F_{MAC} над разрешениями ролей обусловлен применением известного алгоритма комбинирования политик Deny-overrides (рис. 2), при котором в случае, если какое-либо решение оценивается как запрещающее или ни одно решение не оценивается как разрешающее (нет соответствующей политики, соответствующей контексту запроса), результатом является запрет на доступ, если же все решения оцениваются как разрешающие, результатом является разрешение.

Так, конфиденциальность информации и контроль информационных потоков обеспечиваются тем, что даже если роли r будут назначены максимально «широкие» разрешения P (в соответствии с моделью RBAC) и при этом будут учтены все возможные атрибуты контекста $UATT$, $OATT$ в части детализированного доступа (в соответствии с моделью ABAC), фактически пользователь сможет воспользоваться только теми разрешениями из множества P' , где выполнены условия: $c(u) \geq c(o)$ для операции чтения и $c(u) = c(o)$ для операции записи.

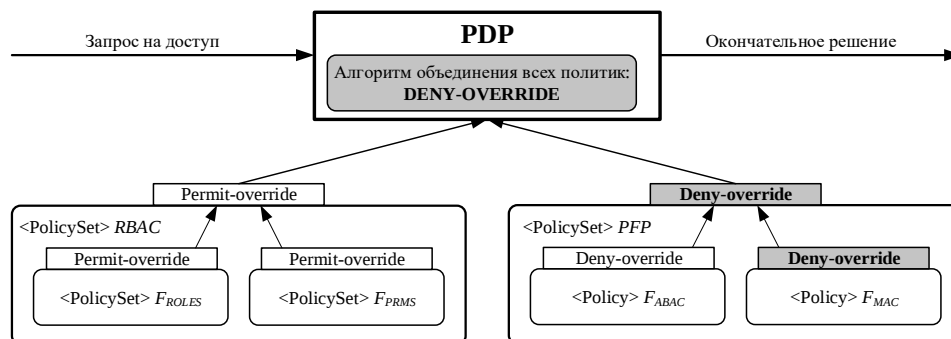


Рис. 2. Схема комбинирования политик безопасности в MAP модели

Важно обратить внимание на то, как в представленной модели обеспечивается выполнение ключевого свойства безопасности модели RBAC, состоящего в том, что для получения доступа к чему-либо субъект должен иметь привязанную роль, имеющую права на этот доступ [20].

Важной частью MAP модели является независимый компонент ABAC, реализованный в виде PFP, что обуславливает необходимость предотвращения проблемы «взрыва правил» в процессе предоставления разрешений на доступ. Политика RBAC и политики набора PFP возвращают разные результаты при обработке запросов на доступ. Тогда как RBAC возвращает только положительные решения Permit (разрешено), PFP используется только для сокращения этих разрешений, соответственно, для сохранения свойства безопасности RBAC, не должно быть функций фильтрации, которые оцениваются как разрешающие. Как и в модели RABAC, это достигается комбинированием всех политик по алгоритму Deny-overrides.

2. Математическое описание модели. Для формального описания MAP модели мы используем следующие множества и функции:

U – множество пользователей;
 S – множество сессий доступа;
 O – множество объектов доступа;
 R – множество ролей;
 $RH \subseteq R \times R$ – иерархия ролей;
 $Cl_{RH}(X)$ – транзитивно-рефлексивное замыкание множества ролей X по RH ;
 P – множество разрешений;
 P' – множество разрешений после фильтрации;
 $OPS = \{read, write\}$ – множество операций;
 $user: S \rightarrow U$, функция принадлежности сессии пользователю;
 $roles: S \rightarrow 2^R$, функция активации ролей в сессии;
 $UA: U \rightarrow 2^R$, функция назначения ролей;
 $PA: R \rightarrow 2^P$, функция назначения разрешений;
 $UA \subseteq U \times R; PA \subseteq R \times P; P \in 2^{(OPS \times O)}; P' \subseteq P$;
 $UATT$ – множество атрибутов пользователя;
 $OATT$ – множество атрибутов объекта;
 $EATT$ – множество атрибутов среды;
 PP – политики фильтрации;
 $L, \leq$ – решетка уровней конфиденциальности;
 $f_{ul}: U \rightarrow L$ – функция уровней доступа пользователей;
 $f_{ol}: O \rightarrow L$ – функция уровней конфиденциальности объектов;
 $f_{sl}: S \rightarrow L$ – функция уровней доступа активируемых сессий;
 $\forall s \in S: f_{sl}(s) \leq f_{ul}(user(s))$;
 $c(u) \in UATT; c(o) \in OATT$ – атрибуты конфиденциальности;
 $\forall att \in UATT \cup OATT \exists Range(att)$ – диапазон атрибута, конечный набор атомарных значений;
 $F_{attType}: UATT \cup OATT \rightarrow \{set, atomic\}$ – функция, задающая тип атрибута (набор или атомарный);
 $F_{filter.p}: P \rightarrow 2^{PP}$, функция фильтрации разрешений, отображение P на множество политик фильтрации разрешений PP ;
 $\forall pp \in PP, pp: U \times R \times P \times 2^{UATT} \times 2^{OATT} \times 2^{EATT} \rightarrow \{t, f\}$;
 $P' = \{p \in P \mid \forall pp \in PP: pp(u, r, p, UATT, OATT, EATT) \neq deny\}$.

Модель формально выражается как детерминированный конечный автомат [21] в виде набора $V = (\Sigma, Q, \phi, q_0, Q_F)$, где Σ и Q – входной алфавит и множество состояний, $q_0 \in Q$ – начальное состояние, $Q_F \subset Q$ – множество конечных состояний, $\phi: Q \times \Sigma \rightarrow Q$ – функция переходов. $\Sigma = \{id, r, p, uatt, oatt, eatt, c(u), c(o)\}$, где $id, c(u), uatt \in UATT$; $c(o), oatt \in OATT$; $eatt \in EATT$, при этом $id, c(u), c(o)$ – обязательные символы каждого входного слова языка $L \subseteq \Sigma^*$, остальные символы – атрибуты контекста. В Q_F входят только два состояния – доступ разрешен (q_t), доступ запрещен (q_f).

Переходы из состояния в состояние (рис. 3) можно описать следующим образом:
 $\varphi(q_0, id(u)) = q_1$ – начало сессии, извлечение R на основе id пользователя;
 $\varphi(q_1, R) = q_2$ – извлечение разрешений P из набора R ;
 $\varphi(q_2, \{P, UATT, OATT, EATT, c(u), c(o)\}) = q_3$ – получение P' путем фильтрации на основе атрибутов безопасности и атрибутов контекста;
 $\varphi(q_3, P') = q_t$ – доступ разрешен;
 $\varphi(q_3, \varepsilon) = q_f$ – отказ в доступе если $P' = \emptyset$.

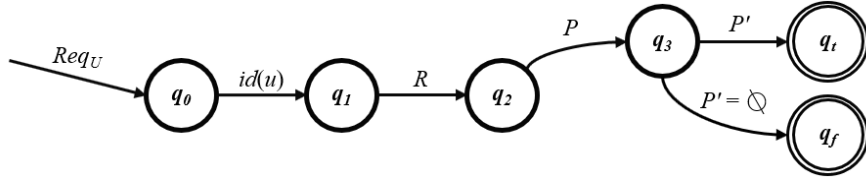


Рис. 3. Диаграмма переходов абстрактного автомата MAP модели

Доказательство безопасности. В рамках MAP модели условия безопасности формулируются через невозможность нарушения требований мандатного управления доступом (запрет на создание явных запрещенных информационных потоков «сверху-вниз»). В отличие от известной модели мандатного ролевого управления доступом [22] выполнение заданных в определении 1 свойств безопасности достигается тем, что исключение запрещенных операций происходит не статически (путем разделения ролей «на чтение» x_read и «на запись» x_write), а динамически, с помощью атрибутивной политики F_{MAC} , которая накладывает дополнительные ограничения на процесс активации сессии s и исключает из ролей (имеющих права и на чтение, и запись) ту или иную операцию согласно правилам политики.

Определение 2. Модель атрибутивного управления доступом на основе ролей соответствует требованиям строгого мандатного управления доступом, когда иерархия на множестве ролей RH в виде отношения частичного порядка « $\leq$ » соответствует требованиям, по которым разрешения на операцию чтения наследуются, если уровень конфиденциальности объекта доступа $c(o_1)$ нижестоящей роли не превышает уровень конфиденциальности объекта доступа $c(o_2)$ вышестоящей роли, разрешения на операцию записи от нижестоящей роли к вышестоящей не наследуются и выполняются ограничения:

- ♦ ограничение функции активации ролей в сессии $roles$: для каждой сессии $s \in S$ выполняется условие $roles(s) \subseteq Cl_{RH}(\{r \mid (user(s), r) \in UA\})$;

- ♦ ограничение функции назначения разрешений ролям PA :

- а) назначение разрешений ролевыми политиками:

$$allow_{RBAC}(u, o, ops) :\Leftrightarrow \exists s \in S: user(s) = u \wedge \exists r \in roles(s): (r, (o, ops)) \in PA;$$

- б) успешное прохождение разрешениями этапов фильтрации PFP (политика F_{MAC}):

$$F_{MAC}(u, o, read) :\Leftrightarrow c(u) \geq c(o);$$

$$F_{MAC}(u, o, write) :\Leftrightarrow c(u) = c(o);$$

- в) получение разрешения пользователем (Deny-overrides):

$$permit(u, o, ops) :\Leftrightarrow allow_{RBAC}(u, o, ops) \wedge F_{MAC}(u, o, ops).$$

В рамках модели дадим определение информационного потока.

Определение 3. Информационный поток от объекта $o \in O$ к объекту $o' \in O$ существует тогда и только тогда, когда существуют роли $r, r' \in R$, сессия s такие, что разрешение на чтение объекта $(o, read) \in PA(r)$, разрешение на запись объекта $(o', write) \in PA(r')$, фактически разрешенные операции пользователя $u = user(s)$: $permit(u, o, read) \wedge permit(u, o', write)$ и роли r, r' активированы в одной сессии s .

Обоснуем, что в предлагаемой модели, соответствующей требованиям строгого мандатного управления доступом, невозможна реализация запрещенных информационных потоков от объектов с высоким уровнем конфиденциальности к объектам с низким уровнем конфиденциальности.

Теорема 1. Если модель атрибутивного управления доступом на основе ролей соответствует требованиям строгого мандатного управления доступом, то в ней для любых объектов $o, o' \in O$ таких, что $c(o) > c(o')$, невозможно возникновение информационного потока от o к o' .

Доказательство. От противного. Пусть существуют объекты $o, o' \in O$ такие, что $c(o) > c(o')$ и возможно возникновение информационного потока от o к o' . По определению 3 существуют роли $r, r' \in R$, сессия s такие, что разрешение на чтение объекта $(o, read) \in PA(r)$, разрешение на запись объекта $(o', write) \in PA(r')$ и роли r, r' активированы в одной сессии s . Следовательно, атрибутивной политикой фильтрации F_{MAC} реализуются требования строгого мандатного управления доступом и выполняются условия: уровень доступа пользователя $c(u) \geq c(o)$ и одновременно $c(u) = c(o')$. Отсюда $c(o) \leq c(o')$.

Противоречие. Если $c(o) > c(o')$, то ни один пользователь в рамках одной сессии не может одновременно получить разрешение на чтение o и запись o' . Теорема доказана.

3. Реализация модели. Закономерным итогом любого исследования в области технических наук является практическое внедрение его результатов. Для демонстрации возможности такого внедрения в данном разделе мы сосредоточились на том, как MAP модель может быть реализована на базе архитектуры XACML. В частности, мы приводим пример реализации, по аналогии с тем, как это сделано в работе [14] и предлагаем вариант кода политики безопасности, порождаемой моделью.

Пример. Защищаемой системой в рамках примера может считаться любая государственная или частная КМИС, имеющая в своем составе множество разнородных информационных систем, где одновременно обрабатывается информация различных уровней конфиденциальности и, ввиду особенностей функционирования, для обеспечения гибкого, детализированного доступа пользователей к ресурсам с учетом множества факторов (контекста) реализуется эталонная архитектура управления доступом ABAC/XACML в соответствии со стандартами [7, 8].

Сценарий доступа. Во множестве информационных систем, входящих в КМИС, реализуется множество совместных проектов, требующих информационного взаимодействия пользователей. Каждый пользователь и объект в КМИС имеют соответствующие метки безопасности в виде атрибутов уровня доступа, атрибутов уровня конфиденциальности соответственно. Для каждой из информационных систем существует отдельная роль с правами на доступ строго к объектам этой системы, а также существует единая гостевая роль для информационных систем, имеющая права на доступ строго к объектам в рамках совместных проектов (единая роль – результат решения проблемы «взрыва ролей» с помощью модели RBAC, в классическом RBAC потребовалось бы создание и назначение отдельных ролей под каждый проект и под каждое новое условие). Пользователю должен быть предоставлен доступ к объекту в рамках совместного проекта, реализуемого в сторонней информационной системе из состава КМИС, с учетом контекста (путем использования всевозможных атрибутивных политик, например: доступ к объекту в рамках совместного проекта может быть одобрен только в рабочее время; запрос должен быть сделан с любого устройства, входящего в состав КМИС; пользователю могут быть разрешены операции только с объектами, относящимся к проектам, в которых он участвует и др.) и необходимости соблюдения требований строгого мандатного управления доступом.

На начальном этапе в точку соблюдения политики РЕР подсистемы авторизации системы управления доступом КМИС (рис. 4) поступает запрос $Z = (id, c(u), UATT)$ аутентифицированного пользователя u на доступ к объекту o , где id – идентификатор пользователя, $c(u)$ – обязательный атрибут уровня доступа пользователя, $UATT$ – конечное множество атрибутов пользователя, сформированное в зависимости от условий ($c(u) \in UATT, c(o) \in OATT$, но для наглядности примера выделяются в отдельные атрибуты, поскольку являющиеся обязательными элементами каждого запроса на доступ независимо от контекста).

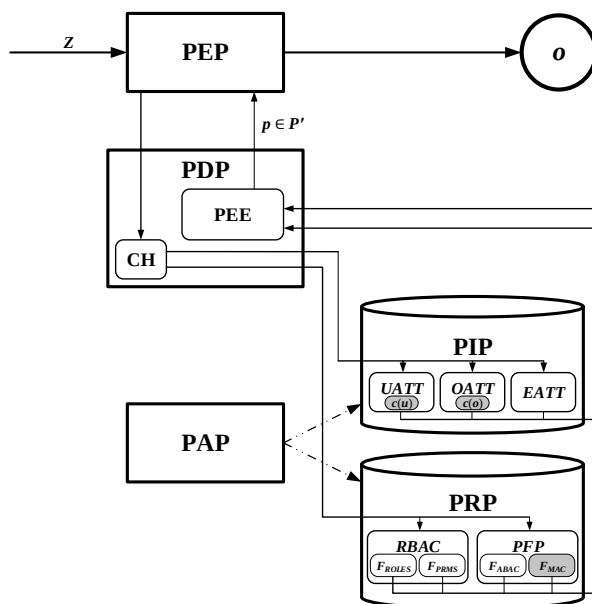


Рис. 4. Реализация MAP модели в подсистеме авторизации системы управления доступом, архитектура которой определяется стандартом XACML

РЕР перенаправляет запрос Z в точку принятия решения по политике PDP, где модуль обработки контекста СН преобразует запрос Z формата соответствующей операционной среды в запрос формата доступа XACML, на основании которого PDP извлекает: из точки информирования по политике PIP, которая является логическим хранилищем атрибутов пользователей, объектов и среды, соответствующие значения атрибутов; из точки извлечения политик PRP, которая является логическим хранилищем политик безопасности, соответствующие политики и наборы политик, алгоритмы их комбинирования.

Далее, на основе полученных атрибутов пользователя, объекта и среды (EATT), политик, наборов политик и алгоритмов их комбинирования в модуле оценки политик РЕЕ точки принятия решения по политике PDP происходит вычисление итогового решения о доступе путем выполнения следующих операций:

- 1) используя политику F_{ROLES} набора политик назначения ролей и разрешений (политика RBAC), пользователю u на основе его id назначается набор ролей R , $u \rightarrow R$;
- 2) используя политику F_{PRMS} набора политик назначения ролей и разрешений (политика RBAC), ролям R назначается множество разрешений P , $R \rightarrow P$;
- 3) используя набор политик PFP (политика ABAC), из множества разрешений P удаляются разрешения, нарушающие правила, сформированные на основе функций фильтрации $F_1, F_2, \dots, F_n$ соответствующих политик и значений атрибутов UATT, OATT и EATT в соответствии с контекстом запроса;
- 4) используя политику F_{MAC} (политика MAC), из множества разрешений P удаляются разрешения, нарушающие правила, согласно которым для операции чтения должно выполняться условие $c(u) \geq c(o)$, для операции записи должно выполняться условие $c(u) = c(o)$, и, таким образом, вычисляется окончательное доступное множество разрешений P' , $P \rightarrow P'$.

Далее решение о доступе (либо об отказе в доступе) направляется из PDP обратно в РЕР с целью его исполнения: при отрицательном решении РЕР блокирует запрос на доступ Z , при положительном решении обеспечивает выполнение операции с объектом.

Благодаря реализации представленной модели на базе архитектуры XACML, в защищаемой системе обеспечивается выполнение ss-свойства и строгого *-свойства политикой F_{MAC} одновременно, путем динамической проверки точкой принятия решения по политике (PDP) каждого перехода между состояниями в защищаемой системе в момент

запроса на этот переход, неизбежно поступающего в точку соблюдения политики (PEP). Так как атрибут уровня конфиденциальности объекта $s(o)$ в разработанной модели является обязательным для каждого объекта защищаемой системы, функции фильтрации политики F_{MAC} будут вызываться для проверки *каждой* операции из P . Таким образом, в окончательное доступное множество разрешений P' войдут только те разрешения, которые соответствуют принципам мандатного разграничения доступа.

Профиль XACML. Учитывая все вышесказанное, мы реализовали набор политик мандатного атрибутивно-ролевого управления доступом в коде XACML 3.0. В настоящий момент этот набор политик прошел процедуру государственной регистрации в виде программы для ЭВМ, Роспатентом выдано соответствующее свидетельство [23].

Заключение. В этой работе мы развили подход RABAC, предложив новую мандатную атрибутивно-ролевую модель управления доступом. На основе анализа открытых источников, это первая модель, объединяющая механизмы MAC, RBAC и ABAC в единую композицию с сохранением их преимуществ. Мировую новизну нашего решения также подтверждает получение патента на изобретение Российской Федерации [24].

Результаты исследования имеют логико-семантический и формальный характер и ориентированы, прежде всего, на подтверждение свойств безопасности модели, которое открывает путь к более широкому внедрению систем управления доступом на основе атрибутов. Использование модели позволит оперативно управлять политиками безопасности при необходимости предоставления детализированного доступа и одновременно обеспечить контроль информационных потоков в защищаемой системе при обработке в ней информации различных уровней конфиденциальности. По нашему мнению, применение модели может быть особенно востребовано в КМИС с высокими требованиями к безопасности, разворачиваемых в средах виртуализации (облако ресурсов и виртуальных машин).

Дальнейшая задача исследования заключается в совершенствовании разработанной модели как в части обеспечения конфиденциальности, так и в части повышения доступности информации, как ключевых качеств функционирования системы управления доступом, и дополнении ее профиля XACML. Также важным направлением будущей работы является полное формализованное (машиночитаемое) описание MAP модели с помощью формальных методов, позволяющих строить абстрактные автоматные модели, а также ее верификация.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Богаченко Н.Ф. Анализ проблем управления разграничением доступа в крупномасштабных информационных системах // Математические структуры и моделирование. – 2018. – № 2 (46). – С. 135-152. – DOI: 10.25513/2222-8772.2018.2.135-152. – EDN: UZQFKY.
2. Harrison M., Ruzzo W., Ullman J. Protection in operating systems // Communication of ACM. – 1976. – 19 (8). – P. 461-471. – DOI: 10.1145/360303.360333.
3. Bell D.E., LaPadula L.J. Secure Computer Systems: Mathematical Foundations // MITRE. – 1973. – Technical Report 2547, Vol. I. – P. 33.
4. Sandhu R., Ferraiolo D., Kuhn R. The NIST model for role-based access control: towards a unified standard // Proceedings of the Fifth ACM Workshop on Role-Based Access Control (RBAC '00). – ACM, Berlin, 2000. – P. 47-63. – DOI: 10.1145/344287.344301.
5. Девянин П.Н. Ролевая ДП-модель управления доступом и информационными потоками в операционных системах семейства Linux // Прикладная дискретная математика. – 2012. – № 1 (15). – С. 69-90. – EDN: OXBJYD.
6. Патент РФ 2525481, МПК G06F 21/62. Способ обеспечения безопасности информационных потоков в защищенных информационных системах с мандатным и ролевым управлением доступом / П.Н. Девянин. – Заявка 2012146550/08, заявлено 01.11.2012, опубликовано 20.08.2014.
7. Ferraiolo D., Chandramouli R., Hu V., Kuhn R. A Comparison of Attribute Based Access Control (ABAC) Standards for Data Service Applications. – NIST.SP.800-178. – Gaithersburg, MD, 2016. – 68 p.
8. ГОСТ Р 59383-2021. Информационные технологии. Методы и средства обеспечения безопасности. Основы управления доступом. – М.: Стандартинформ, 2021. – 35 с.
9. Kuhn D.R., Coyne E.J., Weil T.R. Adding Attributes to Role-Based Access Control // IEEE Computer. – 2010. – 43 (6). – P. 79-81. – DOI: 10.1109/MC.2010.155.

10. Руководящий документ. Автоматизированные системы. Защита от несанкционированного доступа к информации. Классификация автоматизированных систем и требования по защите информации: утв. решением председателя ГТК при Президенте РФ от 30.03.1992. – ФСТЭК России, 1992. – 25 с.
11. Требования о защите информации, не составляющей государственную тайну, содержащейся в государственных информационных системах: утв. Приказом ФСТЭК России от 11.02.2013 № 17. – ФСТЭК России, 2013. – 42 с.
12. Jin X., Krishnan R., Sandhu R. A unified attribute-based access control model covering DAC, MAC and RBAC // *Data and Applications Security and Privacy XXVI (DBSec 2012). Lecture Notes in Computer Science.* – Vol. 7371. – Springer, Berlin, Heidelberg, 2012. – P. 41-55. – DOI: 10.1007/978-3-642-31540-4_4.
13. Kerr L., Alves-Foss J. Combining mandatory and attribute-based access control // *49th Hawaii International Conference on System Sciences (HICSS).* – IEEE, Koloa, HI, 2016. – P. 2616-2623. – DOI: 10.1109/HICSS.2016.328.
14. Jin X., Krishnan R., Sandhu R. RABAC: role-centric attribute-based access control // *Computer Network Security (MMM-ACNS 2012). Lecture Notes in Computer Science.* – Vol. 7531. – Springer, Berlin, Heidelberg, 2012. – P. 84-96. – DOI: 10.1007/978-3-642-33704-8_8.
15. Rajpoot Q.M., Jensen C.D., Krishnan R. Attributes Enhanced Role-Based Access Control Model // *Proceedings of the 12th International Conference on Trust, Privacy and Security in Digital Business (TrustBus'15).* – Springer, Cham, 2015. – P. 3-17. – DOI: 10.1007/978-3-319-22906-5_1.
16. Qi H., Di X., Li J. Formal definition and analysis of access control model based on role and attribute // *Journal of information security and applications.* – 2018. – Vol. 43. – P. 53-60. – DOI: 10.1016/j.jisa.2018.09.001.
17. Houhou O., Bitam S., Hamida A. HYARBAC: a new hybrid access control model for cloud computing // *International Journal of Computing and Digital Systems.* – 2024. – Vol. 15, No. 1. – P. 403-414. – DOI: 10.12785/ijcds/150131.
18. Shahraki A.S., Rudolph C., Alavizadeh H. et al. Securing cross-domain data access with decentralized attribute-based access control // *Ad Hoc Networks.* – 2025. – Vol. 173. – 103807. – P. 1-15. – DOI: 10.1016/j.adhoc.2025.103807.
19. Девянин П.Н., Кулямин В.В., Петренко А.К. [и др.]. Интеграция мандатного и ролевого управления доступом и мандатного контроля целостности в верифицированной иерархической модели безопасности операционной системы // *Тр. ИСП РАН.* – 2020. – Т. 32, № 1. – С. 7-26. – DOI: 10.15514/ISPRAS-2020-32(1)-1. – EDN: WSVFME.
20. Девянин П.Н. Модели безопасности компьютерных систем. Управление доступом и информационными потоками. – 3-е изд. – М.: Горячая линия – Телеком, 2024. – 352 с.
21. Хопкрофт Д.Э., Мотвани Р., Ульман Д. Введение в теорию автоматов, языков и вычислений. – М.: Издательский дом Вильямс, 2008. – 528 с.
22. Sandhu R. Role-based access control // *Advances in Computers.* – 1998. – 46. – P. 237-286. – DOI: 10.1016/S0065-2458(08)60206-5.
23. Свидетельство о государственной регистрации программы для ЭВМ № 2025687752, Российская Федерация. Программный модуль набора политик мандатного атрибутивно-ролевого управления доступом в крупномасштабных информационных системах / Д.О. Ларин. – Заявлено 09.10.2025, опубликовано 15.10.2025. – EDN: RNTBUM.
24. Патент на изобретение РФ 2847174, МПК G06F 21/62. Способ обеспечения конфиденциальности информации в гетерогенных крупномасштабных распределенных информационных системах с атрибутивным управлением доступом / Д.О. Ларин, Р.И. Захарченко, С.А. Диченко. – Заявка 2025112755, заявлено 12.05.2025, опубликовано 29.09.2025. – EDN: EFYGAL.

REFERENCES

1. Bogachenko N.F. Analiz problem upravleniya razgranicheniem dostupa v krupnomasshtabnykh informatsionnykh sistemakh [Analysis of problems of access control management in large-scale information systems], *Matematicheskie struktury i modelirovanie* [Mathematical Structures and Modeling], 2018, No. 2 (46), pp. 135-152. DOI: 10.25513/2222-8772.2018.2.135-152. EDN: UZQFKY.
2. Harrison M., Ruzzo W., Ullman J. Protection in operating systems, *Communication of ACM*, 1976, 19 (8), pp. 461-471. DOI: 10.1145/360303.360333.
3. Bell D.E., LaPadula L.J. Secure Computer Systems: Mathematical Foundations, *MITRE*, 1973. Technical Report 2547, Vol. I, pp. 33.
4. Sandhu R., Ferraiolo D., Kuhn R. The NIST model for role-based access control: towards a unified standard, *Proceedings of the Fifth ACM Workshop on Role-Based Access Control (RBAC '00)*. ACM, Berlin, 2000, pp. 47-63. DOI: 10.1145/344287.344301.

5. Devyanin P.N. Rolevaya DP-model' upravleniya dostupom i informatsionnymi potokami v operatsionnykh sistemakh semeystva Linux [Role-based DP-model for managing access and information flows in Linux operating systems], *Prikladnaya diskretnaya matematika* [Applied Discrete Mathematics], 2012, No. 1 (15), pp. 69-90. EDN: OXBJYD.
6. Devyanin P.N. Patent RF 2525481, MPK G06F 21/62. Sposob obespecheniya bezopasnosti informatsionnykh potokov v zashchishchennykh informatsionnykh sistemakh s mandatnym i rolevym upravleniem dostupom [Patent RF 2525481, IPC G06F 21/62. Method of securing information flow in secure information systems with mandatory and role-based access control]. Application 2012146550/08, filed November 1, 2012, published August 20, 2014.
7. Ferraiolo D., Chandramouli R., Hu V., Kuhn R. A Comparison of Attribute Based Access Control (ABAC) Standards for Data Service Applications. NIST.SP.800-178. Gaithersburg, MD, 2016, 68 p.
8. GOST R 59383-2021. Informatsionnye tekhnologii. Metody i sredstva obespecheniya bezopasnosti. Osnovy upravleniya dostupom [GOST R 59383-2021. Information technologies. Methods and tools of ensuring security. Fundamentals of access control]. Moscow: Standartinform, 2021, 35 p.
9. Kuhn D.R., Coyne E.J., Weil T.R. Adding Attributes to Role-Based Access Control, *IEEE Computer*, 2010, 43 (6), pp. 79-81. DOI: 10.1109/MC.2010.155.
10. Rukovod'yashchiy dokument. Avtomatizirovannyye sistemy. Zashchita ot nesanktsionirovannogo dostupa k informatsii. Klassifikatsiya avtomatizirovannykh sistem i trebovaniya po zashchite informatsii: utv. resheniem predsedatelya GTK pri Prezidente RF ot 30.03.1992 [Guidance document. Automated systems. Protection against unauthorized access to information. Classification of automated systems and information protection requirements: approved by the decision of the Chairman of the State Customs Committee under the President of the Russian Federation dated March 30, 1992]. FSTEK Rossii, 1992, 25 p.
11. Requirements for the protection of information that does not constitute a state secret, contained in state information systems: approved. by Order of the FSTEC of Russia dated 11.02.2013 No. 17 [Requirements for the protection of information that does not constitute a state secret contained in state information systems: approved by Order of the Federal Service for Technical and Export Control of Russia dated February 11, 2013, No. 17]. FSTEC of Russia, 2013, 42 p.
12. Jin X., Krishnan R., Sandhu R. A unified attribute-based access control model covering DAC, MAC and RBAC, *Data and Applications Security and Privacy XXVI (DBSec 2012). Lecture Notes in Computer Science*, Vol. 7371. Springer, Berlin, Heidelberg, 2012, pp. 41-55. DOI: 10.1007/978-3-642-31540-4_4.
13. Kerr L., Alves-Foss J. Combining mandatory and attribute-based access control, *49th Hawaii International Conference on System Sciences (HICSS)*. IEEE, Koloa, HI, 2016, pp. 2616-2623. DOI: 10.1109/HICSS.2016.328.
14. Jin X., Krishnan R., Sandhu R. RABAC: role-centric attribute-based access control, *Computer Network Security (MMM-ACNS 2012). Lecture Notes in Computer Science*, Vol. 7531. Springer, Berlin, Heidelberg, 2012, pp. 84-96. DOI: 10.1007/978-3-642-33704-8_8.
15. Rajpoot Q.M., Jensen C.D., Krishnan R. Attributes Enhanced Role-Based Access Control Model, *Proceedings of the 12th International Conference on Trust, Privacy and Security in Digital Business (TrustBus'15)*. Springer, Cham, 2015, pp. 3-17. DOI: 10.1007/978-3-319-22906-5_1.
16. Qi H., Di X., Li J. Formal definition and analysis of access control model based on role and attribute, *Journal of information security and applications*, 2018, Vol. 43, pp. 53-60. DOI: 10.1016/j.jisa.2018.09.001.
17. Houhou O., Bitam S., Hamida A. HYARBAC: a new hybrid access control model for cloud computing, *International Journal of Computing and Digital Systems*, 2024, Vol. 15, No. 1, pp. 403-414. DOI: 10.12785/ijcds/150131.
18. Shahraki A.S., Rudolph C., Alavizadeh H. et al. Securing cross-domain data access with decentralized attribute-based access control, *Ad Hoc Networks*, 2025, Vol. 173, 103807, pp. 1-15. DOI: 10.1016/j.adhoc.2025.103807.
19. Devyanin P.N., Kulyamin V.V., Petrenko A.K. [i dr.]. Integratsiya mandatnogo i rolevogo upravleniya dostupom i mandatnogo kontrolya tselostnosti v verifitsirovannoy ierarkhicheskoy modeli bezopasnosti operatsionnoy sistemy [Integrating RBAC, MIC, and MLS in Verified Hierarchical Security Model for Operating System], *Tr. ISP RAN* [Proceedings of the Institute for System Programming of the RAS], 2020, Vol. 32, No. 1, pp. 7-26. DOI: 10.15514/ISPRAS-2020-32(1)-1. EDN: WSVFME.
20. Devyanin P.N. Modeli bezopasnosti komp'yuternykh sistem. Upravlenie dostupom i informatsionnymi potokami [Computer system security models. Access and information flow control]. 3rd ed. Moscow: Goryachaya liniya – Telekom, 2024, 352 p.

21. *Khopkroft D.E., Motvani R., Ul'man D.* Vvedenie v teoriyu avtomatov, yazykov i vychisleniy [Introduction to automata theory, languages, and computation]. Moscow: Izdatel'skiy dom Vil'yams, 2008, 528 p.
22. *Sandhu R.* Role-based access control, *Advances in Computers*, 1998, 46, pp. 237-286. DOI: 10.1016/S0065-2458(08)60206-5.
23. *Larin D.O.* Svidetel'stvo o gosudarstvennoy registratsii programmy dlya EVM № 2025687752, Rossiyskaya Federatsiya. Programmnyy modul' nabora politik mandatnogo atributivno-rolevogo upravleniya dostupom v krupnomasshtabnykh informatsionnykh sistemakh [Certificate of state registration of a computer program No. 2025687752 RF. A software module for a set of mandatory role-centric attribute-based access control policies in large-scale information systems]. Announced October 9, 2025, published October 15, 2025. EDN: RNTBUM.
24. *Larin D.O., Zakharchenko R.I., Dichenko S.A.* Patent na izobretenie RF 2847174, MPK G06F 21/62. Sposob obespecheniya konfidentsial'nosti informatsii v geterogennykh krupnomasshtabnykh raspredelennykh informatsionnykh sistemakh s atributivnym upravleniem dostupom [Patent RF 2847174, IPC G06F 21/62. Method for ensuring information confidentiality in heterogeneous large-scale distributed information systems with attribute-based access control]. Application 2025112755, submitted May 12, 2025, published September 29, 2025. EDN: EFYGAL.

Ларин Даниил Олегович – Краснодарское высшее военное училище им. генерала армии С.М. Штеменко; e-mail: jwlll@bk.ru; г. Краснодар, Россия; тел.: +79999830810; адъюнкт; ORCID 0009-0006-4943-0273; WoS ResearcherID LOS-0856-2024.

Захарченко Роман Иванович – Краснодарское высшее военное училище им. генерала армии С.М. Штеменко; e-mail: romanzakharchenko@yandex.ru; г. Краснодар, Россия; тел.: +78612581030; д.т.н.; доцент; начальник кафедры.

Диченко Сергей Александрович – Краснодарское высшее военное училище им. генерала армии С.М. Штеменко; e-mail: dichenko.sa@yandex.ru; г. Краснодар, Россия; тел.: +78612683805; д.т.н.; начальник управления научно-исследовательского центра.

Larin Daniel Olegovich – General of the Army S.M. Shtemenko Krasnodar Higher Military School; e-mail: jwlll@bk.ru; Krasnodar, Russia; phone: +79999830810; Graduate Student (PhD); ORCID 0009-0006-4943-0273; WoS ResearcherID LOS-0856-2024.

Zaharchenko Roman Ivanovich – General of the Army S.M. Shtemenko Krasnodar Higher Military School; e-mail: romanzakharchenko@yandex.ru; Krasnodar, Russia; phone: +78612581030; dr. eng. of sc.; associate professor; head of Department.

Dichenko Sergej Aleksandrovich – General of the Army S.M. Shtemenko Krasnodar Higher Military School; e-mail: dichenko.sa@yandex.ru; Krasnodar, Russia; phone: +78612683805; dr. eng. of sc.; head of the Research Center Directorate.

Раздел III. Электроника, нанотехнологии и приборостроение

УДК 621.396

DOI 10.18522/2311-3103-2026-1-192-208

С.С. Быбин, Н.П. Дунаев, С.В. Кузьмин, А.Н. Морозов

АНАЛИЗ ПРИЧИН ВОЗНИКНОВЕНИЯ ОШИБОК В АМПЛИТУДНО-ФАЗОВОМ РАСПРЕДЕЛЕНИИ ЛИНЕЙНЫХ ФАЗИРОВАННЫХ АНТЕННЫХ РЕШЕТОК И МЕТОДИКА ИХ УМЕНЬШЕНИЯ

Для использования фазированной антенной решётки в режиме диаграммоформирования необходимо установить определённое амплитудно-фазовое распределение на входах излучающих элементов. Амплитудные и фазовые ошибки искажают диаграмму направленности. В работе анализируются источники ошибок в амплитудно-фазовом распределении фазированных антенных решёток, включая паразитные фазовые сдвиги, нелинейность усилительных трактов, температурную нестабильность и взаимные электромагнитные связи между элементами. Описаны три методики калибровки и настройки фазированных антенных решёток, основанные на прямых измерениях коэффициентов передачи в ближней зоне и последующем вычислении вектора воздействий с использованием обратных и псевдообратных матриц взаимных связей, что обеспечивает системный подход к устранению ошибок. Для получения начальных значений предварительно проводилась калибровка каждого канала по закрытому тракту с помощью векторного анализатора цепей. Методика 1 реализует корректировку по множеству точек в пространстве и одному набору состояний каждого канала. Для повышения устойчивости решения методика 2 использует регуляризацию элементов матрицы взаимных связей на основе дополнительного набора измеренных состояний каждого канала. Методика 3 позволяет построить математическую модель конкретной реализации фазированной антенной решётки на основе измерений при фиксированном состоянии каналов, что обеспечивает формирование произвольного амплитудно-фазового состояния без повторных измерений. Проведена экспериментальная настройка восьмизлементной эквидистантной линейной фазированной антенной решётки. Модули аттенюатора/фазовращателя решётки основаны на отладочных платах фазовращателя РЕ44820 и аттенюатора РЕ4302 и управляются с помощью микроконтроллера для автоматического изменения фаз и амплитуд. Измерения проводились автоматически на стенде ближнего поля в безэховой экранированной камере с помощью векторного анализатора цепей. Представлены результаты калибровки, построены матрицы взаимных связей и сформированы диаграммы направленности, подтверждающие работоспособность предложенных подходов. Поскольку экспериментальная решётка является малоэлементной, результаты применения рассмотренных методик сравниваются с результатами настройки в дальней зоне, выполненной с использованием эволюционного алгоритма.

ФАР; калибровка ФАР; определение АФР; измерение характеристик ФАР; эксплуатация ФАР; эволюционные алгоритмы; генетический алгоритм.

S.S. Bybin, N.P. Dunaev, S.V. Kuzmin, A.N. Morozov

ANALYSIS OF THE CAUSES OF ERRORS IN THE AMPLITUDE-PHASE DISTRIBUTION OF LINEAR PHASED ANTENNA ARRAYS AND METHODS FOR THEIR REDUCTION

To use a phased array antenna in the beamforming mode, it is necessary to establish a certain amplitude-phase distribution at the inputs of the emitting elements. Amplitude and phase errors distort the radiation pattern. The paper analyzes the sources of errors in the amplitude-phase distribution of phased antenna arrays, including parasitic phase shifts, nonlinear amplification paths, temperature instability and mutual electromagnetic coupling between the elements. Three methods of calibration and adjustment

of phased antenna arrays are described, based on direct measurements of the transmission coefficients in the near zone and the subsequent calculation of the impact vector using inverse and pseudo-inverse matrices of mutual connections, which provides a systematic approach to error elimination. To obtain the initial values, each channel was pre-calibrated along a closed path using a vector network analyzer. Technique 1 implements correction for a set of points in space and one set of states of each channel. To increase the stability of the solution, method 2 uses the regularization of the elements of the matrix of interconnections based on an additional set of measured states of each channel. Method 3 makes it possible to construct a mathematical model of a specific implementation of a phased array antenna based on measurements with a fixed channel state, which ensures the formation of an arbitrary amplitude-phase state without repeated measurements. An experimental setup of an eight-element equidistant linear phased array antenna was carried out. The lattice attenuator/phase shifter modules are based on the PE44820 phase shifter and PE4302 attenuator debugging boards and are controlled by a microcontroller to automatically change phases and amplitudes. The measurements were carried out automatically on a near-field stand in an anechoic shielded chamber using a vector network analyzer. Calibration results are presented, matrices of mutual relationships are constructed and radiation patterns are formed, confirming the operability of the proposed approaches. Since the experimental array is low-element, the results of applying the considered techniques are compared with the results of tuning in the far zone performed using an evolutionary algorithm.

Phased antenna array; calibration of phased antenna array; determination of amplitude-phase distribution; measurement of characteristics of phased antenna array; operation of phased antenna array; evolutionary algorithms, genetic algorithm.

Введение. На этапе проектирования фазированной антенной решётки (ФАР) с последовательным формированием диаграмм направленности (ДН) различной формы и/или направления, производят учёт взаимных связей методом парциальных диаграмм [1]. В результате решения задачи синтеза ДН, получают требуемые амплитудно-фазовые распределения (АФР) [2, 3]. При помощи статистической теории антенн [4, 5], а чаще, с помощью имитационного моделирования [6], исследуют влияние ошибок в АФР на ДН. Указанные ошибки носят как стохастический характер (уход электрической длины кабелей при изменении температуры...), так и детерминированный характер (дискретность фазовращателей и аттенюаторов (Ф/А)...). По результатам исследования проводят корректировку требуемого АФР. Например, если требуется низкий уровень боковых лепестков, то амплитудное распределение делают ещё более спадающим к краям. Для получения ДН с уровнем боковых лепестков минус 25 дБ, берут распределение, обеспечивающее уровень минус 30 дБ.

После конструирования и изготовления антенной решётки необходимо сформировать требуемое АФР на входах излучателей с учётом ошибок и взаимных связей, или, для краткости, настроить ФАР. В начале, антенную решётку калибруют. Под калибровкой антенной решётки понимают определение характеристик каналов антенны, которые в дальнейшем используются в алгоритмах диаграммоформирования, т.е. определение начальных электрических длин и амплитуд каналов или начальных коэффициентов передачи [7].

Причины, влияющие на ошибки в АФР. *Паразитное изменение фазы, при изменении амплитуды и амплитуды, при изменении фазы.* Наиболее сильно эта особенность работы проявляется в ФАР и АФАР, где используются дискретные Ф/А. В ЦАФАР можно точнее устанавливать фазу и амплитуду в каналах, но, из-за наличия усилителя мощности, в них так же меняется амплитуда и фаза, поскольку АЧХ и ФЧХ усилителя мощности зависит от комплексного входного/выходного сопротивления.

Дискретное изменение амплитуды и фазы в функциональных элементах антенной решётки. Как правило, современные интегральные Ф/А являются дискретными. Поэтому становится невозможно установить произвольную амплитуду и фазу в каналах, что должно учитываться при проектировании.

Технологический разброс параметров материалов и конструкций и влияние нелинейности трактов АФАР. При калибровке нужно выровнять электрические длины всех активных модулей АФАР. Если эти модули соединены с излучателями кабельными сборками, то калибровку нужно производить совместно с кабельными сборками, уложенными в штатные жгуты в составе решётки. Желательно применять хорошо экранированные и фазостабильные кабели.

В антенных решётках, зачастую, применяется спадающее к краям амплитудное распределение. Т.е. в центре решётки УМ могут быть ближе к нелинейному режиму. Нелинейность характеристик проявляется в том, что не только амплитуда, но и фаза начинают меняться нелинейно.

Диаграммообразующая схема (ДОС) может иметь значительную длину. Диэлектрическая проницаемость качественного высокочастотного материала задаётся с точностью $\pm 0,05$. Разность фаз между двумя линиями длиной 400 мм с крайними значениями диэлектрической проницаемости $3,5 - 0,05$ и $3,5 + 0,05$ составит 54° на частоте 5,0 ГГц. ДОС из разных партий могут иметь различную электрическую длину в пределах $\pm 27^\circ$.

Сочленение активного модуля с ДОС внесёт свой вклад в фазовую ошибку. Известно, что при очередном сочленении-расчленении разъёма, даже с помощью динамометрического ключа, фаза коэффициента передачи может меняться на несколько градусов.

Изменение характеристик каналов антенной решётки при изменении внешних воздействующих факторов. По результатам исследования АО «Микран» (Томск) [8] для фазостабильного метрологического кабеля длиной 1 м, при изменении температуры от 25 до 65°C , изменение фазы коэффициента передачи составляет 10° , а амплитуды 0,1 дБ на частоте 5,0 ГГц. При изменении температуры от 0 до минус 50°C изменение фазы составляет минус 6° , а амплитуды — минус 0,15 дБ на частоте 5,0 ГГц. Все остальные узлы фидерного тракта обладают своими характеристиками с точки зрения стабильности амплитуды и фазы при измерении температуры. Они должны быть получены в ходе измерений в камере тепла-холода.

Предполагается, что все каналы антенной решётки подвержены влиянию изменения температуры одинаково и это не сильно сказывается на интегральных характеристиках антенной решётки. Но при спадающем к краям амплитудном распределении центральные элементы решётки могут нагреваться больше, чем периферийные, что приведёт к фазовым ошибкам.

Взаимное влияние элементов антенной решётки. Особенностью работы антенной решётки является наличие взаимных связей между излучающими элементами [9]. Основным механизмом возникновения взаимной связи между излучающими элементами решётки связан с рассеянием поля каждого элемента остальными.

Кроме того, дополнительный вклад в искажение амплитудно-фазового распределения даёт взаимная связь через фидерную часть антенной решётки. На величину просачивающегося паразитного сигнала влияет конечная развязка в делителях мощности, экранировка коаксиальных кабелей, модулей и блоков, наводки по цепям питания.

Ситуация усложняется, если при изменении АФР меняются сопротивления нагрузок излучающих элементов. Например, при изменении состояния фазовращателей меняется не только фаза, но и их входное и выходное сопротивления. Поэтому стараются стабилизировать выходное сопротивление активных модулей за счёт применения ферритовых изделий и/или схем с развязанными направленными ответвителями. В противном случае процесс калибровки и настройки может не сойтись.

Цель исследования:

Исследовать методики уменьшения ошибок амплитудно-фазового распределения в фазированных антенных решётках.

Задачи исследования:

- ◆ Выбрать (предложить) методики калибровки.
- ◆ Реализовать алгоритмы калибровки.
- ◆ Провести калибровку антенной решётки.
- ◆ Провести измерения диаграмм направленности для оценки работоспособности методик и алгоритмов калибровки.

Объект и предмет исследования:

Объектом исследования является амплитудно-фазовое распределение в фазированных антенных решётках.

Предметом исследования являются методики минимизации ошибок в амплитудно-фазовом распределении.

Методики уменьшения ошибок в АФР. После изготовления ФАР необходимо, путём калибровки и настройки, превратить её в инструмент для диаграммоформирования. Причём, желательно получить возможность формирования не только фиксированных ДН в рабочих режимах, но и адаптировать АФР при отказах элементов решётки в условиях эксплуатации. Для этого требуется построить математическую модель конкретной реализации ФАР. Рассмотрим методики, позволяющие осуществить настройку антенной решётки по результатам измерений в ближней зоне.

Линейная ФАР содержит N элементов с номерами $n = 1 \dots N$. Зонд, находящийся в ближней зоне ФАР, может устанавливаться в положениях $p = 1 \dots J$. Далее будем считать, что решётка эквидистантная, как и положение зонда. Число кодов управления $c = 1 \dots I$.

В дальнейшем встречаются следующие множества:

A. Синтезированное АФР

Содержит АФР, полученное в результате решения задачи синтеза, которое необходимо сформировать на входах излучателей. $n(\mathbf{A}) = N$.

B. Потенциально реализуемые коэффициенты передачи

Содержит коэффициенты передачи, которые могут быть потенциально реализованы в функциональных элементах антенной решётки.

Например, для рассматриваемой далее антенной решётки это будут все комплексные числа с амплитудой от 0 дБ до минус 31,5 дБ с шагом 0,5 дБ (6 бит) и фазой от 0° до 360° с шагом $1,4^\circ$ (8 бит).

D. Коды управления

Содержит коды управления функциональными элементами. $n(\mathbf{D}) = n(\mathbf{B})$.

Gn. Коэффициенты передачи в канале антенной решётки

Содержит коэффициенты передачи между входом решётки и выходом на излучатель в конкретном канале при изменении всех кодов управления. Амплитуда нормируется на максимум в канале с минимальной амплитудой по решётке. $n(\mathbf{Gn}) = n(\mathbf{D})$.

V. Поле в ближней зоне

Содержит коэффициенты передачи между ФАР и зондом в ближней зоне. $n(\mathbf{V}) = J$.

Провести прямые измерения возможно только для **Gn** и **V**.

Основной целью настройки является нахождение кодов управления (**Ad**) для формирования синтезированного АФР на входах излучателей **Ad** $\subset$ **D**.

Для всех положений зонда можно записать систему уравнений [10]:

$$\begin{pmatrix} SW_{11} & \dots & SW_{n1} & \dots & SW_{N1} \\ & & \dots & & \\ SW_{1p} & \dots & SW_{np} & \dots & SW_{Np} \\ & & \dots & & \\ SW_{1J} & \dots & SW_{nJ} & \dots & SW_{NJ} \end{pmatrix} \cdot \begin{pmatrix} f_{c1} \\ \dots \\ f_{cn} \\ \dots \\ f_{cN} \end{pmatrix} = \begin{pmatrix} v_1 \\ \dots \\ v_p \\ \dots \\ v_J \end{pmatrix} \quad (1)$$

или

$$\mathbf{SW} \cdot \mathbf{F} = \mathbf{V},$$

где **SW** – матрица взаимных связей; **F** – вектор воздействия или АФР на входах излучателей; **V** – вектор, элементы которого пропорциональны напряжённости поля в местах установки зонда.

Взяв обратную [10, 11] или псевдообратную [12] матрицу от матрицы взаимных связей можно найти вектор воздействий:

$$\mathbf{F} = \begin{cases} \mathbf{SW}^{-1} \cdot \mathbf{V} & (J = N); \\ \mathbf{SW}^+ \cdot \mathbf{V} & (J > N). \end{cases} \quad (2)$$

Вектор **V** зависит от вектора воздействий и непосредственно измеряется как коэффициент передачи между входом ФАР и выходом зонда.

Вектор F , не может быть измерен в составе решётки. Он либо вычисляется из (2), либо считается заданным – получается в ходе проектирования при решении задачи синтеза ДН. Результаты калибровки помогают скорректировать вектор воздействия.

Матрица SW в общем случае зависит от F , но в первом приближении этим можно пренебречь. Особенно при рассмотрении АФАР и ЦАФАР, в которых излучатели нагружены на постоянные нагрузки. В [10] приводится методика получения SW через многократные измерения коэффициентов передачи при разных F . Также возможно рассчитать SW с учётом комплексных парциальных ДН излучателей и ДН зонда в ближней зоне.

Фактически, при настройке антенных решёток, как указано в [13], решается задача об извлечении количественной информации о физических закономерностях из данных об их косвенных проявлениях. Находится F из V . Такая задача является обратной. Кроме того, поскольку V измеряется с ошибками и решение может быть неединственным и неустойчивым по исходным данным, задача является некорректной и требует регуляризации.

Далее будут рассмотрены несколько методик, одна из которых является обобщением [10] и [12].

Методика настройки №1 ($J = 20 \div 40 \cdot N$) [12]

1. Калибровка по закрытому тракту

В результате калибровки каналов ФАР, содержащих функциональные элементы, получается связь между B , D и Gn .

1.1. Излучатели отсоединяются, измеряются коэффициенты передачи между входом решётки и выходами модулей Ф/А при изменении амплитуд и фаз. Из-за паразитных изменений получаются, например, разные фазы при изменении амплитуды и неизменном коде на фазовращателе.

1.2. Амплитуды в каждом канале решётки нормируются на максимум канала с минимальной амплитудой (относительно всей решетки). Иначе, во всех каналах, кроме одного, появятся нереализуемые состояния. К сожалению, это уменьшит эквивалентную изотропно-излучаемую мощность. В случае изначально спадающего к краям амплитудного распределения методика усложняется и должна быть доработана. При этом, как правило, каналы с максимальной амплитудой ставятся в центре решётки.

1.3. Находятся расстояния, на комплексной плоскости, от потенциально реализуемых коэффициентов передачи до измеренных. Выбирается код, соответствующий минимальному расстоянию. В результате, например, получается, что идеальному состоянию минус 1.0 дБ, 30° в канале соответствует код «минус 0.5 дБ, 45°». При этом, реальное значение коэффициента передачи в закрытом тракте минус 1.3 дБ, 33°.

2. Измерение SW

Зонд устанавливается в положение p и измеряются коэффициенты передачи между входом решётки и выходом зонда. При этом, активный канал находится в начальном состоянии 0.0 дБ, 0.0°. Все остальные каналы – с минимально возможной амплитудой. Фактически они играют роль согласованной нагрузки. В таком случае элемент матрицы взаимных связей будет равен коэффициенту передачи. В виде матрицы алгоритм записывается следующим образом:

$$\begin{pmatrix} 1 & \dots & 0 & \dots & 0 \\ \vdots & & & & \\ 0 & \dots & 1 & \dots & 0 \\ \vdots & & & & \\ 0 & \dots & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} SW_{11} & \dots & SW_{1p} & \dots & SW_{1J} \\ \vdots & & \vdots & & \vdots \\ SW_{n1} & \dots & SW_{np} & \dots & SW_{nJ} \\ \vdots & & \vdots & & \vdots \\ SW_{N1} & \dots & SW_{Np} & \dots & SW_{NJ} \end{pmatrix} = \begin{pmatrix} v_{11} & \dots & v_{1p} & \dots & v_{1J} \\ \vdots & & \vdots & & \vdots \\ v_{n1} & \dots & v_{np} & \dots & v_{nJ} \\ \vdots & & \vdots & & \vdots \\ v_{N1} & \dots & v_{Np} & \dots & v_{NJ} \end{pmatrix} \quad (3)$$

или

$$F \cdot SW = V,$$

$$SW = F^{-1} \cdot V$$

3. Обращение $SW \rightarrow SW^+$

4. Измерение V

Производится при заданном АФР, полученном в результате решения задачи синтеза. Коды в каналах устанавливаются исходя из идеальных амплитуд и фаз.

5. Получение $F \rightarrow F = SW^+ \cdot V$

Получаются значения, не совпадающие ни с идеальными значениями, ни с коэффициентами передачи, полученными в результате калибровки.

6. Корректировка F

Например, в канале необходимо было получить точку минус 0.5 дБ, 11° . Из-за описанных ошибок и взаимных связей в антенной решётке, на входе излучателя получается значение элемента вектора воздействия минус 0.9 дБ, 48° . Далее, фактически, проводим новую калибровку, повторяем п.1.3. Ищем точку на комплексной плоскости, ближайшую к точке минус 0.1 дБ, минус 26° и соответствующий ей код управления. Если при этом появляются нереализуемые состояния (амплитуда больше 0 дБ), то в п.1.2 придётся ещё опуститься по амплитуде.

7. Повторение п.п. 4-6.

Повторить указанные пункты можно несколько раз.

При данной методике настройки, матрица взаимных связей SW связывает АФР на входах излучателей с измеряемым полем в ближней зоне (не обязательно на линии), что позволяет провести диагностику решётки в полевых условиях, если вращать антенну или зонд вокруг антенны [12].

Методика настройки №2 ($J = 20 \div 40 \cdot N$)

1. Калибровка по закрытому тракту

2. Измерение SW

Зонд устанавливается напротив первого канала. На первый Ф/А последовательно подаются все коды управления (или их часть) и проводятся измерения коэффициентов передачи. Остальные Ф/А остаются в произвольном состоянии ($c = \tau$). Затем, первый Ф/А устанавливается в состояние согласованной нагрузки, а коды управления подаются на второй Ф/А. Остальные Ф/А по прежнему остаются в состоянии согласованной нагрузки. После того, как доходят до последнего Ф/А, зонд устанавливают напротив второго канала и начинают процедуру опять с первого Ф/А. Этот процесс для любого p и первого Ф/А описывается системой уравнений:

$$\begin{pmatrix} f_{11} & \dots & f_{\tau n} & \dots & f_{\tau N} \\ & & \dots & & \\ f_{c1} & \dots & f_{\tau n} & \dots & f_{\tau N} \\ & & \dots & & \\ f_{I1} & \dots & f_{\tau n} & \dots & f_{\tau N} \end{pmatrix} \cdot \begin{pmatrix} SW_{1p} \\ \dots \\ SW_{np} \\ \dots \\ SW_{Np} \end{pmatrix} = \begin{pmatrix} v_{11p} \\ \dots \\ v_{c1p} \\ \dots \\ v_{I1p} \end{pmatrix} \quad (4)$$

или

$$F_1 \cdot SW_p = V_{1p},$$

где

F_1 – матрица воздействий при изменении кодов управления в первом канале;

SW_p – строка матрицы взаимных связей;

V_{1p} – вектор, элементы которого пропорциональны напряжённости поля в местах установки зонда при изменении кодов управления в первом канале.

Для любого n и p (4) записывается, как:

$$F_n \cdot SW_p = V_{np}$$

Заметим, что уравнение (4) можно переписать в виде:

$$\begin{pmatrix} f_{11} & 1 \\ \dots & \\ f_{c1} & 1 \\ \dots & \\ f_{I1} & 1 \end{pmatrix} \cdot \begin{pmatrix} SW_{1p} \\ g_{1p} \end{pmatrix} = \begin{pmatrix} v_{11p} \\ \dots \\ v_{c1p} \\ \dots \\ v_{I1p} \end{pmatrix} \quad (5)$$

или, уже для любого n ,

$$\mathbf{f}_n \cdot \mathbf{sg}_{np} = \mathbf{V}_{np},$$

где $\mathbf{g}_{np}$ – константа.

Таким образом, с помощью псевдообратных матриц $\mathbf{f}_n^+$, можно найти все члены матрицы $\mathbf{SW}$.

В простейшем случае, при большом затухании в тракте, $f_\tau = 0$, по аналогии с (3), (5) можно записать для $n = 1$ в виде:

$$\begin{pmatrix} f_{11} \cdots 0 \cdots 0 \\ \vdots \\ f_{c1} \cdots 0 \cdots 0 \\ \vdots \\ f_{l1} \cdots 0 \cdots 0 \end{pmatrix} \begin{pmatrix} sw_{11} \cdots sw_{1p} \cdots sw_{1J} \\ \vdots \\ 0 \cdots 0 \cdots 0 \\ \vdots \\ 0 \cdots 0 \cdots 0 \end{pmatrix} = \begin{pmatrix} v_{111} \cdots v_{11p} \cdots v_{11J} \\ \vdots \\ v_{c11} \cdots v_{c1p} \cdots v_{c1J} \\ \vdots \\ v_{l11} \cdots v_{l1p} \cdots v_{l1J} \end{pmatrix}, \quad (6)$$

для $n = 2$:

$$\begin{pmatrix} 0 \ f_{12} \cdots 0 \cdots 0 \\ \vdots \\ 0 \ f_{c2} \cdots 0 \cdots 0 \\ \vdots \\ 0 \ f_{l2} \cdots 0 \cdots 0 \end{pmatrix} \begin{pmatrix} 0 \cdots 0 \cdots 0 \\ sw_{21} \cdots sw_{2p} \cdots sw_{2J} \\ \vdots \\ 0 \cdots 0 \cdots 0 \\ \vdots \\ 0 \cdots 0 \cdots 0 \end{pmatrix} = \begin{pmatrix} v_{121} \cdots v_{12p} \cdots v_{12J} \\ \vdots \\ v_{c21} \cdots v_{c2p} \cdots v_{c2J} \\ \vdots \\ v_{l21} \cdots v_{l2p} \cdots v_{l2J} \end{pmatrix},$$

для остальных n аналогично.

При этом, методика №2 отличается от методики №1 тем, что производится регуляризация элементов матрицы $\mathbf{SW}$.

3. Обращение $\mathbf{SW} \rightarrow \mathbf{SW}^+$

4. Измерение $\mathbf{V}$

Производится при заданном АФР, полученном в результате решения задачи синтеза. Коды в каналах устанавливаются исходя из идеальных амплитуд и фаз.

5. Получение $\mathbf{F} \rightarrow \mathbf{F} = \mathbf{SW}^+ \cdot \mathbf{V}$

6. Корректировка $\mathbf{F}$

7. Повторение п.п. 4-6.

Недостатком представленных методик №1 и №2 является то, что они требуют измерения $\mathbf{V}$ для формирования каждого нового АФР.

Методика настройки №3

Для получения $\mathbf{SW}$ в Методике №2 производится многократное измерение $\mathbf{V}$ при различных $\mathbf{F}$. В частности, получается, что при фиксированном $\mathbf{F}$ измеряется набор $\mathbf{V}$ для всех положений зонда. По аналогии с (6), для $n = 1$, запишем:

$$\begin{pmatrix} f_{11} \cdots 0 \cdots 0 \\ \vdots \\ f_{c1} \cdots 0 \cdots 0 \\ \vdots \\ f_{l1} \cdots 0 \cdots 0 \end{pmatrix} \begin{pmatrix} sw_1 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \sum_{p=1}^J v_{11p} \\ \vdots \\ \sum_{p=1}^J v_{c1p} \\ \vdots \\ \sum_{p=1}^J v_{l1p} \end{pmatrix},$$

откуда

$$\mathbf{SW}_1 = \mathbf{F}_1^+ \cdot \mathbf{V}_1.$$

Для остальных n аналогично.

После нахождения всех элементов матрицы взаимных связей можно записать следующее уравнение:

$$\begin{pmatrix} sw_1 & \dots & 0 & \dots & 0 \\ \vdots & & & & \\ 0 & \dots & sw_n & \dots & 0 \\ \vdots & & & & \\ 0 & \dots & 0 & \dots & sw_N \end{pmatrix} \begin{pmatrix} f_{c1} \\ \vdots \\ f_{cn} \\ \vdots \\ f_{cN} \end{pmatrix} = \begin{pmatrix} f'_{c1} \\ \vdots \\ f'_{cn} \\ \vdots \\ f'_{cN} \end{pmatrix},$$

или

$$SW \cdot F = F', \quad (7)$$

где SW – матрица взаимных связей;

F – вектор воздействий, получаемый из решения задачи синтеза АФР;

F' – вектор воздействий, который необходимо установить на входах излучателей.

Подобный переход от полей, измеренных в ближней зоне, к АФР на входах излучателей [14, 15] позволяет получить математическую модель конкретной реализации антенной решётки. Что, в свою очередь, даёт возможность формировать произвольные АФР и осуществлять адаптацию ДН при выходе из строя части каналов.

Результаты настройки восьмиэлементной линейной ФАР. Настройка и измерения характеристик проводились для восьмиэлементной АФАР на частоте 2 ГГц при $d = 0.5 \cdot \lambda = 75$ мм. Схема измерений и внешний вид решётки представлены на рис. 1.

Антенная решётка включает в себя восемь модулей аттенюатора/фазовращателя. Модули основаны на отладочных платах фазовращателя PE44820 [16] и аттенюатора PE4302 [17]. Запитываются модули параллельно, через делители мощности. Для управления всеми модулями предусмотрено устройство управления (УУ), основанное на отладочной плате STM32H723 [18]. Для измерения коэффициентов передачи применяется векторный анализатор цепей (ВАЦ) ZNB40 (Rohde & Schwarz).

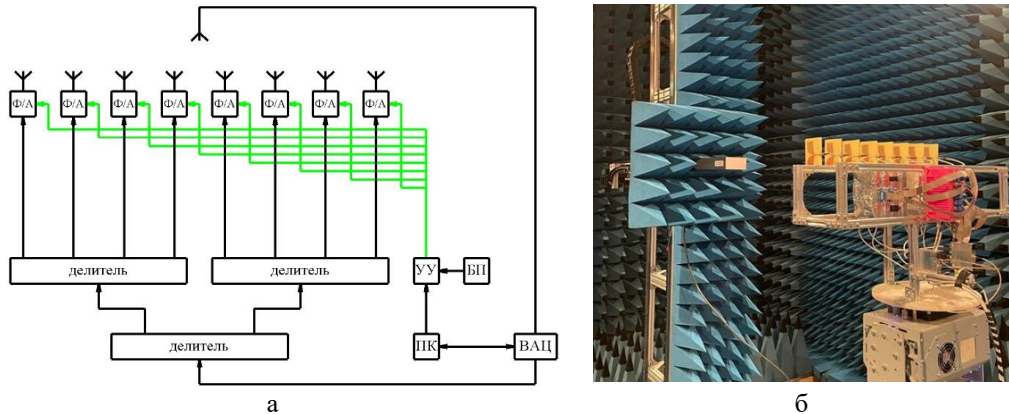


Рис. 1. Схема измерений (а) и внешний вид решётки (б)

Измерения коэффициентов передачи осуществлялось в ближней зоне на расстоянии 200 мм ($\sim 3\lambda$) между излучателями антенной решетки и зондом. Позиционирование, управление модулями АР и запрос данных с ВАЦ выполнялось в автоматическом режиме. Измерения ДН проводились в дальней зоне на расстоянии 5 м.

Для получения коэффициентов передачи в канале антенной решётки каждый модуль поочередно отключался от излучателя и подключался к ВАЦ. В подключенном модуле автоматически перебирались коды управления и результаты измерений записывались в таблицу. Остальные (неподключенные) модули были в режимах начальной фазы и «минус 31,5 дБ», т.е. играли роль согласованных нагрузок.

Полученные коэффициенты передачи нормировались на максимальное значение в канале с минимальной амплитудой. Среди этих значений по критерию минимального расстояния выбирались точки для искомого АФР, в данном случае синфазного равноамплитудного. На рис. 2 представлены результаты для каналов 1-3, цветными линиями обо-

значены коэффициенты передачи по амплитуде от 0.0 дБ до минус 10.0 дБ с шагом 0,5 дБ и по фазе от 0° до $358,6^\circ$ с шагом $1,4^\circ$. Значения приведены в относительных единицах (вольтах, нормированных на максимальное значение минимального по амплитуде канала).

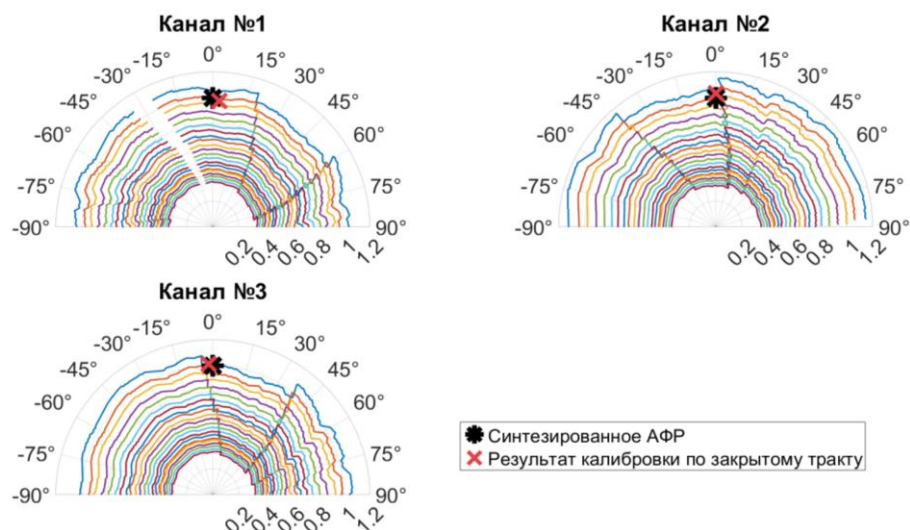


Рис. 2. Коэффициенты передачи в относительных единицах для каналов 1–3 в режимах: по амплитуде от 0 дБ до минус 10 дБ, по фазе от 0° до $358,6^\circ$ с шагом $1,4^\circ$

Физически электрические длины каналов и амплитуды коэффициентов передачи не выравнились. Парциальные ДН несимметричны и отличаются от канала к каналу, что показано на рис. 3. Такая решётка, с большими ошибками по фазе и амплитуде, была создана именно с целью проведения экспериментов по настройке.

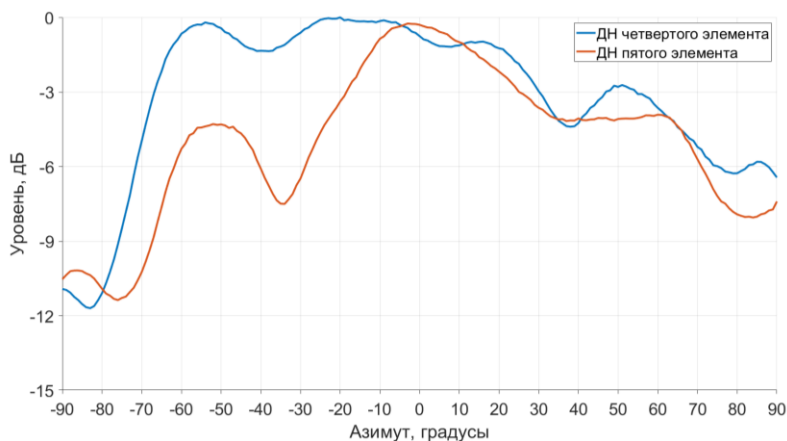


Рис. 3. Парциальные ДН центральных каналов

Если воспользоваться только результатами калибровки по закрытому тракту, то, при синфазном и равноамплитудном АФР получается ДН представленная на рис. 4.

Поскольку антенная решётка содержит всего 8 элементов, она может быть настроена непосредственно перебором АФР при помощи Генетического алгоритма [19]. Решалась задача оптимизации с критерием максимизации S_{21} , что пропорционально коэффициенту усиления. Каждая особь кодировалась двумя вектор-строками, которые содержали коды управления аттенюатором и фазовращателем. Применялась эволюционная стратегия (μ, λ) [19], где из числа детей λ выживают μ . Родители из предыдущих итераций не

доживают. Каждая особь существует одно поколение или одну итерацию. Применение такой стратегии позволило ускорить сходимость с нескольких тысяч итераций до нескольких сотен. Разброс мутации уменьшался с увеличением числа итераций. На рис. 5 представлены три ДН, полученные в результате работы Генетического алгоритма.

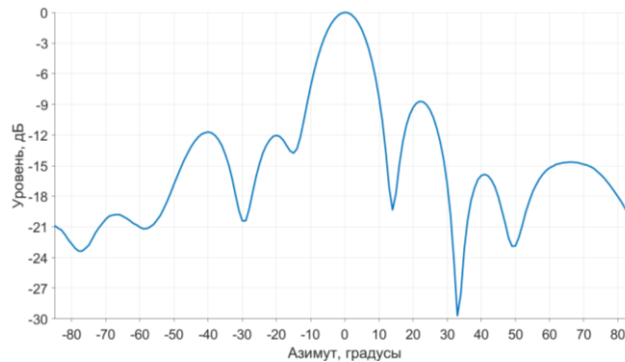


Рис. 4. ДН при синфазном и равноамплитудном АФР, полученном по результатам калибровки по закрытому тракту

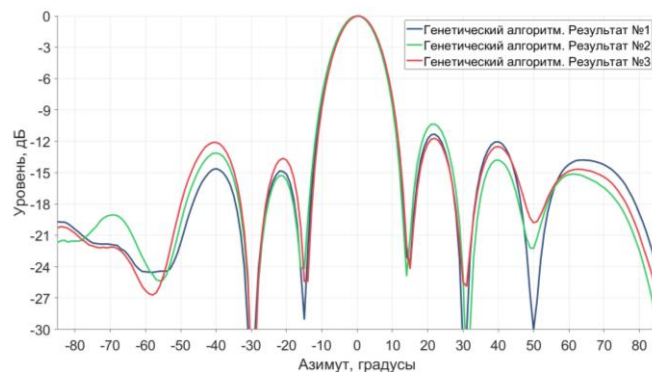


Рис. 5. ДН при максимизации K_u , полученные с применением Генетического алгоритма

Поскольку современное измерительное оборудование позволяет получать ДН достаточно быстро, то оптимизацию можно проводить, не только измеряя коэффициент передачи, но и сравнивая измеренную ДН с заданной огибающей. Соответственно, для малоэлементных ФАР, могут быть получены АФР для формирования ДН специальной формы [20].

Описание алгоритма. По методикам №1 и №2 был написан алгоритм для вычисления АФР на элементах решетки. В случае работы по методике №2 алгоритм учитывал набор кодов в каждом канале и соответствующие измерения в ближней зоне. Методика №1 является частным случаем методики №2 с одним набором кодов управления.

Алгоритм, для получения матриц коэффициентов передачи, в автоматическом режиме перебирает коды и каналы в заданных положениях зонда в пространстве. Полученные данные по вышеописанным формулам преобразовывались в матрицы SW для методик №1 и №2.

Далее, для синфазного равноамплитудного распределения, по результатам калибровки по закрытому тракту измерялся вектор V и решалась система линейных уравнений (2). Для её решения можно применить несколько встроенных функций MATLAB:

- ◆ `pinv`. Вычисление псевдообратной матрицы [21, 22] по методу Мура–Пенроуза;
- ◆ `lsqminnorm`. Метод наименьших квадратов с минимальной нормой решения [23, 22];
- ◆ `Singular Value Decomposition (SVD)`. Сингулярное разложение матрицы [24];
- ◆ `Decomposition`. Разложение в различных режимах работы [25].

Для ускорения вычислений в некоторых методах матрица приводилась к виду ленточной [26] за счет обнуления элементов матрицы меньших $1e-4$. Результаты, полученные с применением различных функций практически совпадают, но требуют различного времени выполнения (рис. 6).

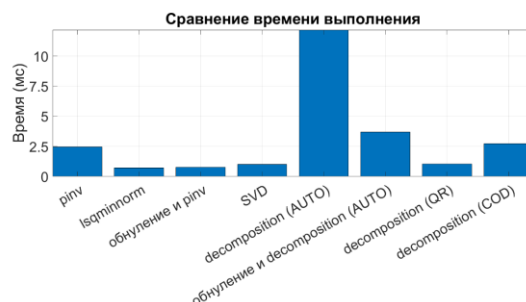


Рис. 6. Сравнение времени решения системы уравнений (2) разными методами

Полученный набор АФР (F) сравнивался с G_p для внесения корректировки. В случае методики №2 вектор V снимался повторно для нового F.

Результаты. В соответствии с описанной Методикой №1 была получена матрица взаимных связей, представленная на рис. 7.

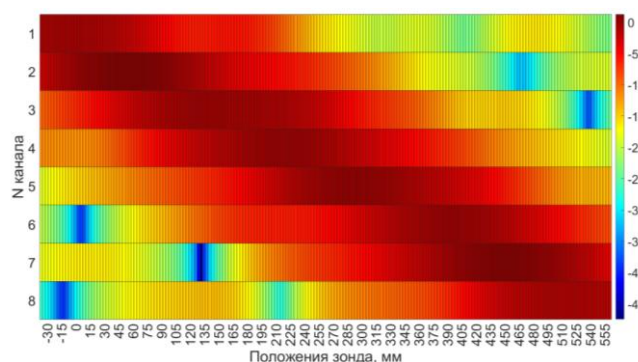


Рис. 7. Матрица взаимных связей [дБ], Методика №1

Результаты применения Методики №1 для формирования синфазного и равноамплитудного АФР показаны на рис. 8.

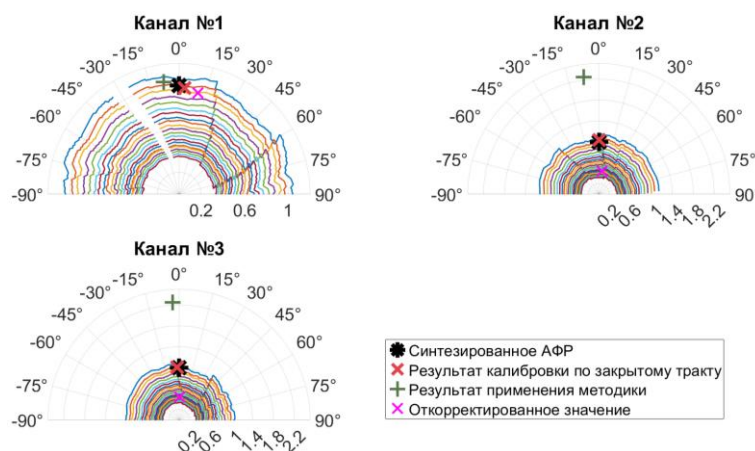


Рис. 8. Результаты применения Методики №1, 1-3 каналы

На рис. 9 представлена ДН ФАР с кодами управления, полученными в результате применения Методики №1.

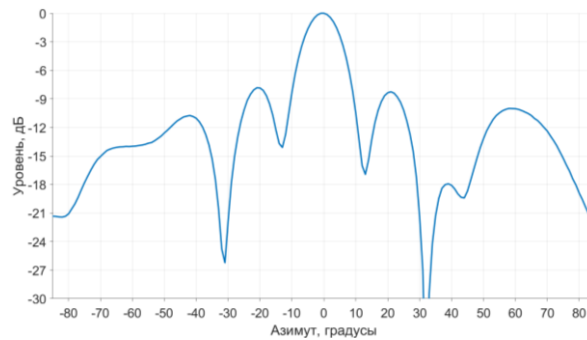


Рис. 9. Диаграмма направленности, полученная по Методике №1

В Методике №1 результат зависит от выбора начального состояния канала. На рис. 10 представлены десять ДН при значениях фазы начального состояния канала от 0° до 101.25° с шагом 11.25° . Поэтому требуется регуляризация элементов матрицы SW .

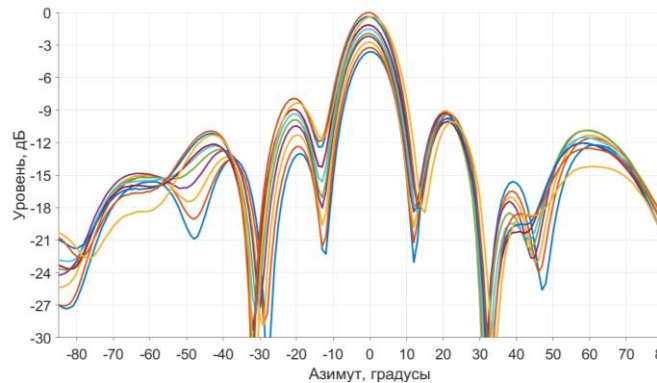


Рис. 10. Диаграмма направленности, полученная по Методике №1 с различными значениями начальной фазы в каналах

В Методике №2, для получения матрицы взаимных связей были выбраны коды управления, соответствующие точкам минус 1.0 дБ по амплитуде и от 0° до 348.75° с шагом 11.25° по фазе. На рис. 11 представлены 32 требуемые точки (красные) и ближайшие к ним коэффициенты передачи в канале антенной решётки (синие точки).

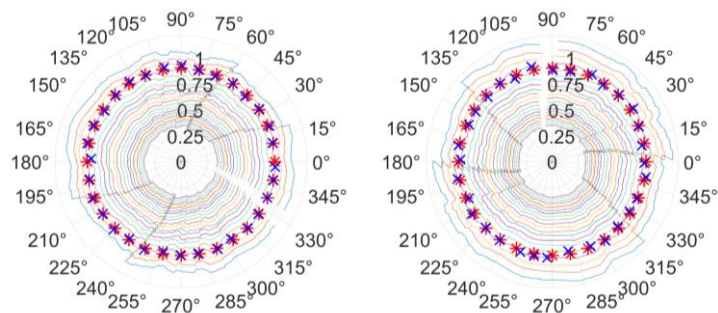


Рис. 11. Расположение исходных и ближайших к ним точек, для измерения матрицы взаимных связей в Методике №2, 1 и 2 каналы

В результате применения Методики №2 была получена матрица взаимных связей и ДН, представленные на рис. 12, 13 соответственно. Алгоритм по методике №2 выполнялся за 5 итераций для последовательного уточнения АФР на элементах решетки.

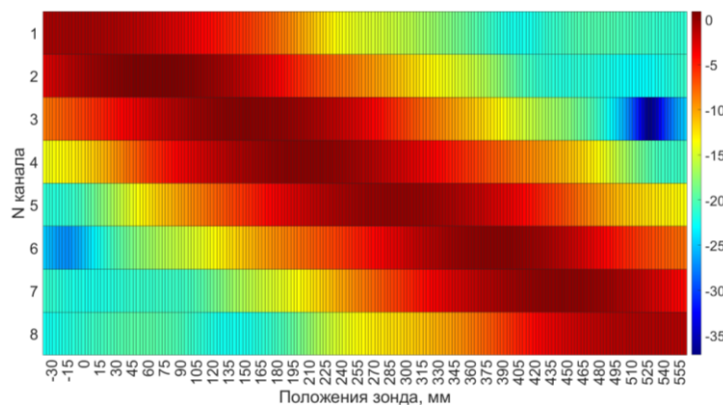


Рис. 12. Матрица взаимных связей [дБ], Методика №2

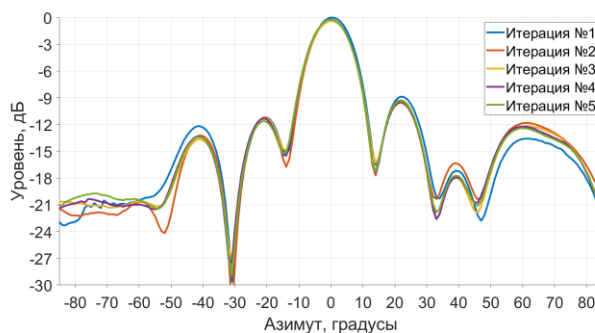


Рис. 13. Диаграммы направленности, полученная по Методике №2

На рис. 14 показан результат применения Методики №3.

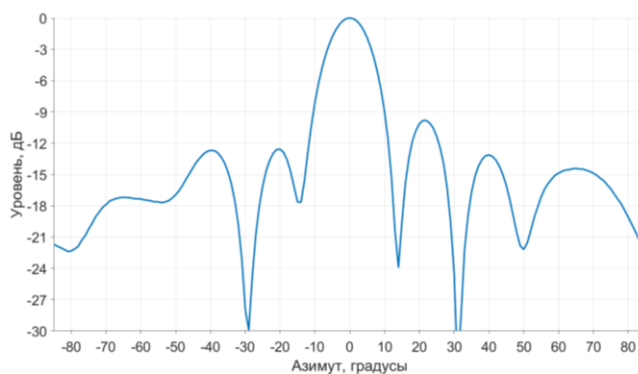


Рис. 14. Диаграмма направленности, полученная по Методике №3

В табл. 1 приведены коды управления аттенюаторами и фазовращателями, которые устанавливались для получения ДН. Коды фазы соответствуют потенциально реализуемым коэффициентам передачи. Коды амплитуды соответствуют шагу аттенюатора 0.5 дБ (код 13 соответствует 6.5 дБ).

Таблица 1

Коды управления, соответствующие результатам работы алгоритмов

	Коды управления															
	Аттенюатор [дБ]								Фазовращатель [градус]							
Калибровка по закрытому тракту	1	0	0	1	0	0	0	0	16	188	163	139	142	16	135	19
Методика №1	1	14	13	15	14	4	11	0	160	68	36	2	7	156	1	155
Методика №2. Ит. 1	1	2	1	2	1	0	0	0	16	187	163	136	140	16	135	20
Методика №2. Ит. 2	2	3	0	3	0	1	1	0	128	35	16	241	255	128	246	135
Методика №2. Ит. 3	2	4	0	3	0	1	1	1	10	171	155	128	135	9	128	15
Методика №2. Ит. 4	2	3	0	3	0	1	1	0	11	172	155	128	135	11	130	16
Методика №2. Ит. 5	2	3	0	3	0	1	1	0	11	172	155	128	135	12	130	16
Методика №3	2	4	2	2	2	2	1	0	155	74	43	16	3	148	18	165
ГА 1 (150 ит.)	0	0	0	0	0	0	0	0	243	143	131	103	083	246	131	15
ГА 2 (150 ит.)	0	0	0	0	0	0	0	0	231	143	119	91	81	242	123	10
ГА 3 (300 ит.)	0	0	0	0	0	0	0	0	231	141	119	91	83	227	121	7

Заключение. В работе исследованы и показаны не все практические аспекты методик и алгоритмов настройки ФАР. В частности, подлежат исследованию вопросы, связанные с:

- ◆ количеством точек установки зонда, что влияет на скорость проведения измерений;
- ◆ шагом решётки, поскольку наличие дифракционных максимумов может повлиять на сходимость методик;
- ◆ количеством точек установки АФР в каналах решётки для получения матрицы взаимных связей, что влияет на скорость проведения измерений;
- ◆ количеством элементов решётки, поскольку при увеличении размеров ФАР появляется возможность формирования ДН специальной формы с низким уровнем боковых лепестков.

Выводы. Не все ошибки в АФР могут быть определены заранее и учтены в ходе проектирования. Калибровка ФАР по закрытому тракту помогает уменьшить уровень ошибок, но не полностью. Особенно трудно учесть взаимные связи с учётом конструкции и технологического разброса характеристик излучающих элементов.

Существуют методики формирования требуемого АФР по результатам изменений в ближней зоне. Для каждой ФАР должны быть предусмотрены ресурсы и время на отработку алгоритмов настройки.

Малоэлементные антенные решётки могут быть настроены по результатам измерений в дальней зоне с применением эволюционных алгоритмов.

Представленные методики формирования АФР на входах излучателей позволяют настроить конкретную реализацию ФАР, что показано в ходе экспериментов с восьми-элементной линейной антенной решёткой.

Методику №3 можно использовать в качестве модели конкретной реализации ФАР при эксплуатации. Например, для восстановления характеристик антенны при выходе из строя некоторых каналов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Андропов А.В., Кузьмин С.В. Методика синтеза диаграмм направленности антенных решеток с произвольным расположением излучающих элементов // Тр. учебных заведений связи. – 2022. – Т. 8, № 2. – С. 15-28.
2. Дмитриева В.В., Коровин К.О., Ликонцев А.Н. Синтез диаграмм направленности антенных решеток для больших углов сканирования с использованием генетического алгоритма // Тр. учебных заведений связи. – 2024. – Т. 10, № 1. – С. 49-57.
3. Brown A.D. Electronically Scanned Arrays MATLAB® Modeling and Simulation. – Boca Raton: CRC Press, 2012. – 229 p. – ISBN. – DOI: 10.1201/b12044.
4. Шифрин Я.С. Вопросы статистической теории антенн. – М.: Сов. радио, 1970. – 383 с.
5. Справочник по антенной технике: в 5 т. Т. 1 / под ред. Л.Д. Бахраха. – М.: Радиотехника, 1997. – 248 с.
6. Кутузов О.И., Татарникова Т.М. Моделирование систем. Имитационный метод: учебник. – СПб.: Лань, 2022. – 223 с.
7. Коротецкий Е.В., Шитиков А.М., Денисенко В.В. Методы калибровки фазированных антенных решеток // Радиотехника. – 2013. – № 5. – С. 95-104.
8. Круглов В.Г., Медников В.В. Исследование отклонения фазы и амплитуды коэффициента отражения фазостабильной кабельной сборки при изменении температуры окружающей среды. – URL: <https://www.micran.ru/about/blog/izmenenie-fazovoy-i-amplitudnoy-stabilnosti-ksf26-pri-izmenenii-t-nn-jn-kvg-j/> (дата обращения: 13.05.2025).
9. Хансен Р.С. Фазированные антенные решетки: пер. с англ. – М.: Техносфера, 2012. – 560 с.
10. Анисимов А.А. и др. Применение матрицы взаимных связей при настройке ФАР // Радиолокация и радиосвязь: Матер. V Всерос. науч.-техн. конф. – М.: ИПЭ РАН, 2011. – С. 3-8.
11. Lei J., Fu G., Yang L., Fu D.-M. Wide band linear printed antenna array with low sidelobe cosecant square-shaped beam pattern // Progress In Electromagnetics Research C. – 2010. – Vol. 15. – P. 233-241. – DOI: 10.2528/PIERC09112702.
12. Шубников В.В., Ривкин М.И. Методика экспресс-анализа РТХ ФАР в эксплуатационных условиях // Вестник воздушно-космической обороны. – 2018. – № 4 (20). – С. 105-108.
13. Головицкий А.П. Обратные задачи экспериментальной физики: практические аспекты: учеб. пособие. – СПб.: СПбГТУ, 2008. – 222 с.
14. Гармаш В.Н., Малакишинов Н.П., Пузанков В.Ф. Численные методы решения некоторых обратных задач восстановления характеристик излучающих систем по измеренным полям в дальней и ближней зонах // Сб. научно-методических статей по прикладной электродинамике. – М.: Высш. шк., 1983. – Вып. 5. – С. 98-130.
15. Хашимов А.Б. Регуляризирующие схемы в задачах восстановления поля излучения антенн по измерениям в ближней // Вестник ЮУрГУ. Серия "Компьютерные технологии, управление, радиоэлектроника". – 2014. – Т. 14, № 2. – С. 70-80.
16. PE44820 – Цифровой 8-битный фазовращатель: datasheet. – URL: <https://psemi.com/pdf/datasheets/pe44820ds.pdf> (дата обращения: 11.07.2025).
17. PE4302 – Цифровой 6-битный аттенуатор: datasheet. – URL: <https://clck.ru/3N4ZjP> (дата обращения: 11.07.2025).
18. STM32H723: datasheet. – URL: <https://clck.ru/3N4ai2> (дата обращения: 11.07.2025).
19. Саймон Д. Алгоритмы эволюционной оптимизации: практическое руководство: пер. с англ. А.В. Логунова. – М.: ДМК Пресс, 2020. – 1002 с.
20. Дунаев Н.П., Кузьмин С.В., Макаров В.А. Опыт применения генетического алгоритма для настройки и калибровки антенной решетки в дальней зоне // Антенны и распространение радио-волн: Сб. докл. Всерос. науч.-техн. конф. – СПб.: СПбГЭТУ "ЛЭТИ", 2023. – С. 75-77.
21. pinv // MATLAB Exponenta: документация. – URL: <https://docs.exponenta.ru/matlab/ref/pinv.html> (дата обращения: 21.08.2024).
22. Strang G. Introduction to Linear Algebra. – 5th ed. – Wellesley: Wellesley-Cambridge Press, 2016. – 574 с. – ISBN 978-0-9802327-7-6.
23. lsqminnorm // MATLAB Exponenta: документация. – URL: <https://docs.exponenta.ru/matlab/ref/lsqminnorm.html#d123e829846> (дата обращения: 21.08.2024).
24. svd // MATLAB Exponenta: документация. – URL: <https://docs.exponenta.ru/matlab/ref/double.svd.html> (дата обращения: 21.08.2024).
25. decomposition // MATLAB Exponenta: документация. – URL: <https://docs.exponenta.ru/matlab/ref/decomposition.html#d123e318933> (дата обращения: 21.08.2024).
26. Голуб Дж., Ван Лоун Ч. Матричные вычисления: пер. с англ. и науч. ред. Б.В. Сайфулина, Ю.А. Кузнецова, В.Б. Степанова. – 3-е изд., стер. – М.: Мир, 1999. – 548 с. – ISBN 5-03-002406-9.

REFERENCES

1. Andropov A.V., Kuz'min S.V. Metodika sinteza diagramm napravlenosti antennykh reshetok s proizvol'nym raspolozheniem izluchayushchikh elementov [Methodology for synthesizing directional patterns of antenna arrays with arbitrary arrangement of radiating elements], *Tr. uchebnykh zavedeniy svyazi* [Proceedings of Communications Educational Institutions], 2022, Vol. 8, No. 2, pp. 15-28.
2. Dmitrieva V.V., Korovin K.O., Likontsev A.N. Sintez diagramm napravlenosti antennykh reshetok dlya bol'shikh uglov skanirvaniya s ispol'zovaniem geneticheskogo algoritma [Synthesis of directional patterns of antenna arrays for large scanning angles using a genetic algorithm], *Tr. uchebnykh zavedeniy svyazi* [Proceedings of Communications Educational Institutions], 2024, Vol. 10, No. 1, pp. 49-57.
3. Brown A.D. Electronically Scanned Arrays MATLAB® Modeling and Simulation. Boca Raton: CRC Press, 20126 229 p. ISBN. DOI: 10.1201/b12044.
4. Shifrin Ya.S. Voprosy statisticheskoy teorii antenn [Questions of the statistical theory of antennas]. Moscow: Sov. radio, 1970, 383 p.
5. Spravochnik po antennoy tekhnike [Handbook of antenna technology]: in 5 vol. Vol. 1, ed. by L.D. Bakhrakha. Moscow: Radiotekhnika, 1997, 248 p.
6. Kutuzov O.I., Tatarnikova T.M. Modelirovanie sistem. Imitatsionnyy metod: uchebnik [Modeling of systems. Simulation method: textbook]. Saint Petersburg.: Lan', 2022, 223 p.
7. Korotetskiy E.V., Shitikov A.M., Denisenko V.V. Metody kalibrovki fazirovannykh antennykh reshetok [Methods for calibrating phased antenna arrays], *Radiotekhnika* [Radio Engineering], 2013, No. 5, pp. 95-104.
8. Kruglov V.G., Mednikov V.V. Issledovanie otkloneniya fazy i amplitudy koeffitsienta otrazheniya fazostabil'noy kabel'noy sborki pri izmenenii temperatury okruzhayushchey sredy [Study of the deviation of the phase and amplitude of the reflection coefficient of a phase-stable cable assembly with changes in the ambient temperature]. Available at: <https://www.micran.ru/about/blog/izmenenie-fazovoy-i-amplitudnoy-stabilnosti-ksf26-pri-izmenenii-t-njn-kvg-j/> (accessed 13 May 2025).
9. Khansen R.S. Fazirovannye antennye reshetki [Phased antenna arrays]: transl. from Engl. Moscow: Tekhnosfera, 2012, 560 p.
10. Anisimov A.A. i dr. Primenenie matritsy vzaimnykh svyazey pri nastroyke FAR [Application of the matrix of mutual connections in tuning a phased array], *Radiolokatsiya i radiosvyaz': Mater. V Vseros. nauch.-tekhn. konf.* [Radar and radio communications: Proceedings of the V All-Russian scientific and technical conference]. Moscow: IRE RAN, 2011, pp. 3-8.
11. Lei J., Fu G., Yang L., Fu D.-M. Wide band linear printed antenna array with low sidelobe cosecant square-shaped beam pattern, *Progress In Electromagnetics Research C*, 2010, Vol. 15, pp. 233-241. DOI: 10.2528/PIERC09112702.
12. Shubnikov V.V., Rivkin M.I. Metodika ekspress-analiza RTKh FAR v ekspluatatsionnykh usloviyakh [Methodology for express analysis of phased array radio-technical characteristics under operating conditions], *Vestnik vozdushno-kosmicheskoy oborony* [Bulletin of Aerospace Defense], 2018, No. 4 (20), pp. 105-108.
13. Golovitskiy A.P. Obratnye zadachi eksperimental'noy fiziki: prakticheskie aspekty: ucheb. posobie [Inverse problems of experimental physics: practical aspects: textbook]. Saint. Petersburg: SPbGTU, 2008, 222 p.
14. Garmash V.N., Malakshinov N.P., Puzankov V.F. Chislennyye metody resheniya nekotorykh obratnykh zadach vosstanovleniya kharakteristik izluchayushchikh sistem po izmerennym polyam v dal'ney i blizhney zonakh [Numerical methods for solving some inverse problems of reconstructing the characteristics of radiating systems from measured fields in the far and near zones], *Sb. nauchno-metodicheskikh statey po prikladnoy elektrodinamike* [Collection of scientific and methodological articles on applied electrodynamics]. Moscow: Vyssh. shk., 1983, Issue 5, pp. 98-130.
15. Khashimov A.B. Regulariziruyushchie skhemy v zadachakh vosstanovleniya polya izlucheniya antenn po izmereniyam v blizhney [Regularizing schemes in problems of antenna radiation field reconstruction from near-field measurements], *Vestnik YuUrGU. Seriya "Komp'yuternye tekhnologii, upravlenie, radioelektronika"* [Bulletin of SUSU. Series "Computer technologies, control, radio electronics"], 2014, Vol. 14, No. 2, pp. 70-80.
16. PE44820 – Tsifrovoy 8-bitnyy fazovrashchator: datasheet [PE44820 – Digital 8-bit phase shifter: datasheet]. Available at: <https://psemi.com/pdf/datasheets/pe44820ds.pdf> (accessed 11 July 2025).
17. PE4302 – Tsifrovoy 6-bitnyy atten'yuator: datasheet [PE4302 – Digital 6-bit Attenuator: datasheet]. Available at: <https://clck.ru/3N4ZjP> (accessed 11 July 2025).
18. STM32H723: datasheet. Available at: <https://clck.ru/3N4ai2> (accessed 11 July 2025).
19. Saymon D. Algoritmy evolyutsionnoy optimizatsii: prakticheskoe rukovodstvo [Evolutionary optimization algorithms: a practical guide]: transl. from Engl. by A.V. Logunova. Moscow: DMK Press, 2020, 1002 p.

20. Dunaev N.P., Kuz'min S.V., Makarov V.A. Opyt primeneniya geneticheskogo algoritma dlya na-stroyki i kalibrovki antennoy reshetki v dal'ney zone [Experience in applying a genetic algorithm to far-field antenna array tuning and calibration], *Antenny i rasprostraneniye radiovoln: Sb. dokl. Vseros. nauch.-tekhn. konf.* [Antennas and Radio Wave Propagation: Collection of Papers from the All-Russian Scientific and Technical Conference]. Saint Petersburg: SPbGETU "LETI", 2023, pp. 75-77.
21. pinv, *MATLAB Exponenta: dokumentatsiya* [MATLAB Exponenta: documentation]. Available at: <https://docs.exponenta.ru/matlab/ref/pinv.html> (accessed 21 August 2024).
22. Strang G. Introduction to Linear Algebra. 5th ed. Wellesley: Wellesley-Cambridge Press, 2016 574 z. ISBN 978-0-9802327-7-6.
23. lsqminnorm, *MATLAB Exponenta: dokumentatsiya* [MATLAB Exponenta: documentation]. Available at: <https://docs.exponenta.ru/matlab/ref/lsqminnorm.html#d123e829846> (accessed 21 August 2024).
24. svd, *MATLAB Exponenta: dokumentatsiya* [MATLAB Exponenta: documentation]. Available at: <https://docs.exponenta.ru/matlab/ref/double.svd.html> (accessed 21 August 2024).
25. Decomposition, *MATLAB Exponenta: dokumentatsiya* [MATLAB Exponenta: documentation]. Available at: <https://docs.exponenta.ru/matlab/ref/decomposition.html#d123e318933> (accessed 21 August 2024).
26. Golub Dzh., Van Loun Ch. Matrichnye vychisleniya [Matrix computations]: transl. from Engl. [and scientific editor] B.V. Sayfulina, Yu.A. Kuznetsova, V.B. Stepanova. 3rd ed. Moscow: Mir, 1999, 548 p. ISBN 5-03-002406-9.

Быбин Сергей Сергеевич – АО "Навигатор"; e-mail: bybins@gmail.com; г. Санкт-Петербург, Россия; тел.: +79219895199; начальник отдела.

Дунаев Николай Павлович – Санкт-Петербургский государственный университет аэрокосмического приборостроения; e-mail: ndunaev@inbox.ru; г. Санкт-Петербург, Россия; тел.: +79026261644; ассистент кафедры радиотехнических систем; инженер АО "Навигатор".

Кузьмин Сергей Викторович – Санкт-Петербургский государственный университет аэрокосмического приборостроения; e-mail: sergey-v-kuzmin@yandex.ru; г. Санкт-Петербург, Россия; тел.: +79119643306; к.ф.-м.н.; доцент кафедры радиотехнических систем; главный специалист АО "Навигатор".

Морозов Александр Николаевич – АО "Навигатор"; e-mail: MorozovAN@navigat.ru; г. Санкт-Петербург, Россия; тел.: +79214204299; начальник научно-исследовательского сектора.

Bybin Sergey Sergeevich – JSC Navigator; e-mail: bybins@gmail.com; St. Petersburg, Russia; phone: +79219895199; head of unit.

Dunaev Nikolay Pavlovich – Saint-Petersburg State University of Aerospace Instrumentation; e-mail: ndunaev@inbox.ru; Saint Petersburg, Russia; phone: +79026261644; assistant of the Department of Radio Engineering Systems; engineer of the JSC Navigator.

Kuzmin Sergey Viktorovich – Saint-Petersburg State University of Aerospace Instrumentation; e-mail: sergey-v-kuzmin@yandex.ru; Saint Petersburg, Russia; phone: +79119643306; cand. of phys. and math. sc.; associate professor of the Department of Radio Engineering Systems; chief specialist of the JSC Navigator.

Morozov Alexander Nikolaevich – JSC Navigator; e-mail: MorozovAN@navigat.ru; Saint Petersburg, Russia; phone: +79214204299; head of the Research sector.

УДК 621.371: 538.574

DOI 10.18522/2311-3103-2026-1-208-218

А.И. Семенихин, Д.В. Семенихина, А.М. Зыкина

МЕТАПОВЕРХНОСТИ С ОРБИТАЛЬНЫМ УГЛОВЫМ МОМЕНТОМ НА ОСНОВЕ ГЕКСАГОНАЛЬНЫХ ЯЧЕЕК ДЛЯ ШИРОКОПОЛОСНОГО СНИЖЕНИЯ РАССЕЯНИЯ

Статья посвящена актуальной проблеме – исследованию возможностей фазового гашения рассеянных электромагнитных волн с помощью тонких непоглощающих метаповерхностей Панчаратна-Берри (ПБ) с генерацией вихревых волн с орбитальным угловым моментом (ОУМ) и спиралевидным фазовым фронтом. Целью работы является проектирование таких метаповерхностей (МП) на основе единичных ячеек гексагональной формы и традиционной квадратной формы и сравнение их характери-

стик рассеяния и возможностей широкополосного фазового гашения рассеяния. Метаповерхности с бездисперсной ПБ-фазой для волн круговой поляризации (КП-волн) состоят из ячеек, в которых углы поворота мета-частиц изменяются по заданному закону. Гексагональная ячейка (типа пчелиная сота) имеет шесть осей симметрии (вместо четырех у квадратной ячейки), что должно обеспечивать меньшее влияние разных углов поворота соседних мета-частиц МП на свойства коэффициентов отражения ячеек. В статье предложена мета-частица в виде перфорированного полоска в гексагональной ячейке, которая эффективно отражает ко-поляризованные КП-волны в полосе от 8,5 до 19,7 ГГц независимо от угла поворота мета-частицы. Спроектированы четыре модели МП из таких ячеек с генерацией ОУМ разного порядка. Симуляция рассеяния КП-волн методом конечных элементов подтвердила, что модели с гексагональными ячейками эффективнее снижают поле обратного рассеяния (свыше 10 дБ) по сравнению с квадратными ячейками (лишь на 8 дБ) в сверхширокой полосе частот от 8,5 до 20,8 ГГц. Обратное рассеяние снижается за счет генерации воронкообразных вихревых волн с модами ОУМ минус первого или плюс третьего порядков и фазовой сингулярности поля на оси вихрей. Полученные результаты могут быть полезны при выборе формы единичных ячеек метаповерхностей, предназначенных для широкополосного гашения рассеяния.

Метаповерхность Панчаратнама-Бэрри; гексагональная единичная ячейка; вихревая волна; орбитальный угловой момент; снижение рассеяния.

A.I. Semenikhin, D.V. Semenikhina, A.M. Zikina

HEXAGONAL CELL-BASED ORBITAL ANGULAR MOMENTUM METASURFACES FOR BROADBAND SCATTERING REDUCTION

The article is devoted to the actual problem to study of the possibilities of phase cancellation of scattered electromagnetic waves using thin non-absorbing Pancharatnam-Berry (PB) metasurfaces with the generation of vortex waves with orbital angular momentum (OAM) and a spiral phase front. The aim of the work is to design such metasurfaces (MS) based on unit cells of hexagonal shape and traditional square shape and to compare their scattering characteristics and the possibilities of broadband phase cancellation of scattering. Metasurfaces with a dispersionless PB-phase for circular polarized waves (CP-waves) consist of cells in which the rotation angles of meta-particles change according to a given law. A hexagonal unit cell (like a honeycomb) has six axes of symmetry (instead of four in a square unit cell), which should provide a smaller influence of different rotation angles of adjacent meta-particles of an MS on the properties of the reflection coefficients of the cells. In this paper, a meta-particle in the form of a perforated patch in a hexagonal cell is proposed, which effectively reflects co-polarized CP-waves in the range from 8.5 to 19.7 GHz regardless of the rotation angle of the meta-particle. Four models of metasurfaces from such cells with generation of OAM of different orders were designed. Simulation of scattering of CP-waves by the finite element method confirmed that models with hexagonal cells reduce the backscattering field more effectively (over 10 dB) compared to square cells (only by 8 dB) in an ultra-wide frequency range from 8.5 to 20.8 GHz. Backscattering is reduced due to the generation of funnel-shaped vortex waves with OAM modes of minus the first or plus the third order and a phase singularity of the field on the vortex axis. The obtained results can be useful in choosing the shape of unit cells of metasurfaces intended for broadband scattering reduction.

Pancharatnam-Berry metasurface; hexagonal cell; vortex wave; orbital angular momentum; scattering reduction.

Введение. Наиболее известными механизмами широкополосного фазового гашения эффективной площади рассеяния блестящих участков металлических тел с помощью непоглощающих метаповерхностей (МП) являются интерференционное гашение (деструктивная интерференция) [1, 2], твист-эффект (поляризационная конверсия) [3, 4], аномальное отражение и диффузное рассеяние [5–8]. Для этого непоглощающие МП проектируют в виде решеток конструктивных элементов (модулей, субрешеток), которые состоят из мета-частиц с кодированными фазами локальных коэффициентов отражения. Интерференционное гашение рассеяния реализуется за счет противофазности соседних шахматно-подобных модулей [1]. Твист-эффект обеспечивает перевод энергии поля на кросс-поляризацию за счет поляризационной конверсии падающих волн линейной поляризации в ортогонально поляризованные волны линейной поляризации. В метаповерхностях с аномальным отражением и диффузным рассеянием применяются более сложные (чаще всего, синтезированные) фазовые профили локальных коэффициентов отражения, улучшающие широкоугольность гашения бистатического рассеяния.

В последние годы в различных частотных диапазонах предложены новые фазовые механизмы управления рассеянием электромагнитных волн и сверхширокополосного гашения рассеяния, основанные на применении метаповерхностей, формирующих бездисперсную фазу Панчаратнама-Берри (ПБ-фазу) для волн круговых поляризаций (КП-волн) [9–14]. Метаповерхности Панчаратнама-Берри (ПБ-метаповерхности) состоят из мета-частиц с поляризационной конверсией другого типа, когда падающая КП-волна преобразуется при отражении без деполяризации в согласно поляризованную КП-волну.

Интенсивно исследуются непоглощающие ПБ-метаповерхности с фазовым механизмом гашения рассеяния за счет генерации вихревых волн (фазовых вихрей) с ненулевым орбитальным угловым моментом (ОУМ), фазовой сингулярностью и обнулением поля на оси вихрей [12–16]. В работе [14] для генерации ОУМ предложены тонкие ПБ-метаповерхности (толщиной порядка $0,1\lambda_n$ относительно длины волны λ_n в подложке на нижней рабочей частоте), состоящие из мета-частиц в виде полосков с вырезами. Гашение кросс-поляризованного поля вихревого рассеяния от МП в широкой полосе 8,5–20,5 ГГц составило лишь 8 дБ.

Традиционно ПБ-метаповерхности проектируются на основе единичных ячеек четырехгранной квадратной формы [12–14]. Однако, известно [17], что периодические решетки, образованные многогранными прямоугольными ячейками, особенно, гексагональные решетки, демонстрируют в ряде приложений лучшие свойства, чем решетки квадратных ячеек. Единичная ячейка в виде правильного шестигранника (типа пчелиная сота) имеет меньшую площадь и шесть осей симметрии вместо четырех у квадратной ячейки. Эти особенности широко используются при проектировании линз, частотно-селективных структур, антенных решеток и метаматериалов, особенно в терагерцовом диапазоне [17–19]. В частности, известно, что использование правильных шестиугольных ячеек повышает плотность упаковки элементов в антеннах на основе МП и позволяет более точно кодировать фазу электромагнитных волн [20] и реализовать более широкую полосу пропускания [21, 22]. Тем не менее, публикаций по метаповерхностям (МП) с ячейками гексагональной формы гораздо меньше.

В настоящей работе с целью улучшения эффективности гашения обратного рассеяния (не менее чем на 10 дБ) рассматриваются ПБ-метаповерхности с широкополосной генерацией ОУМ, состоящие из ячеек гексагональной формы (вместо традиционных квадратных ячеек). Проектируются и сравниваются характеристики рассеяния конструктивных элементов МП (ОУМ-модулей) из гексагональных и квадратных ячеек. Симуляция характеристик рассеяния ко-поляризованных и кросс-поляризованных волн круговой поляризации выполнена в HFSS в случае нормального падения плоских КП-волн.

1. Коэффициенты отражения волн линейных и круговых поляризаций в каналах Флоке квадратной и гексагональной ПБ-ячеек. Геометрия ПБ-ячеек показана на рис. 1. Ячейки содержат экранированную подложку F4b2 ($\epsilon = 2,65$; $\text{tg}\delta = 0,001$) толщиной 3 мм и идентичные мета-частицы толщиной 0,035 мм в виде перфорированных полосков с тремя вырезами, размеры которых для двух видов ячеек приведены на рис. 1. Сторона квадратной ячейки равна 10 мм [14]; диаметр описанной окружности гексагональной – 12,8 мм. Угол поворота β мета-частиц отсчитывается от оси x (рис. 1).

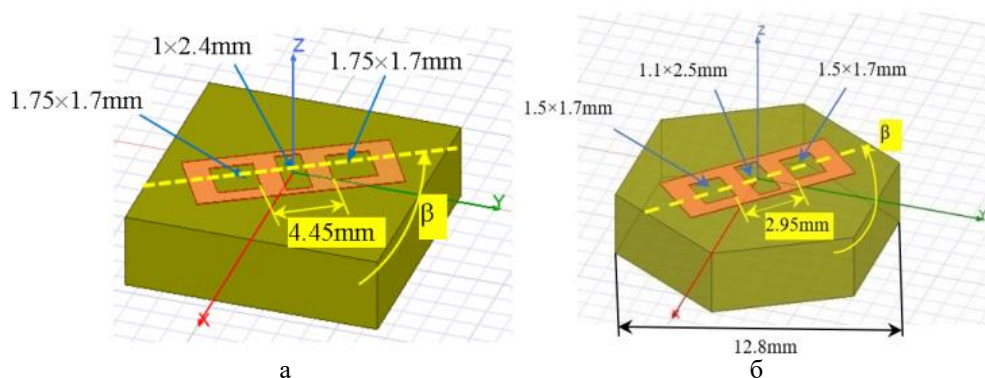


Рис. 1. Единичные ячейки квадратной (а) и гексагональной (б) формы

Для проектирования МП с гексагональными ячейками и симуляции S-параметров МП используем канал Флоке (рис. 2 а,г). В канале Флоке осуществляется анализ бесконечной МП путем наложения периодических граничных условий, как показано на рис. 2,а. Для возбуждения гексагональной единичной ячейки, в канале задан порт Флоке с модами TE_{00} , TM_{00} (рис. 2,б-г). Векторы $\vec{a}$ и $\vec{b}$ порта Флоке (рис. 2,д) с углом 60° между ними являются векторами трансляции узлов двумерной гексагональной решетки. Рис. 2,д поясняет принцип построения бесконечной МП с гексагональной решеткой ячеек с помощью операций трансляции.

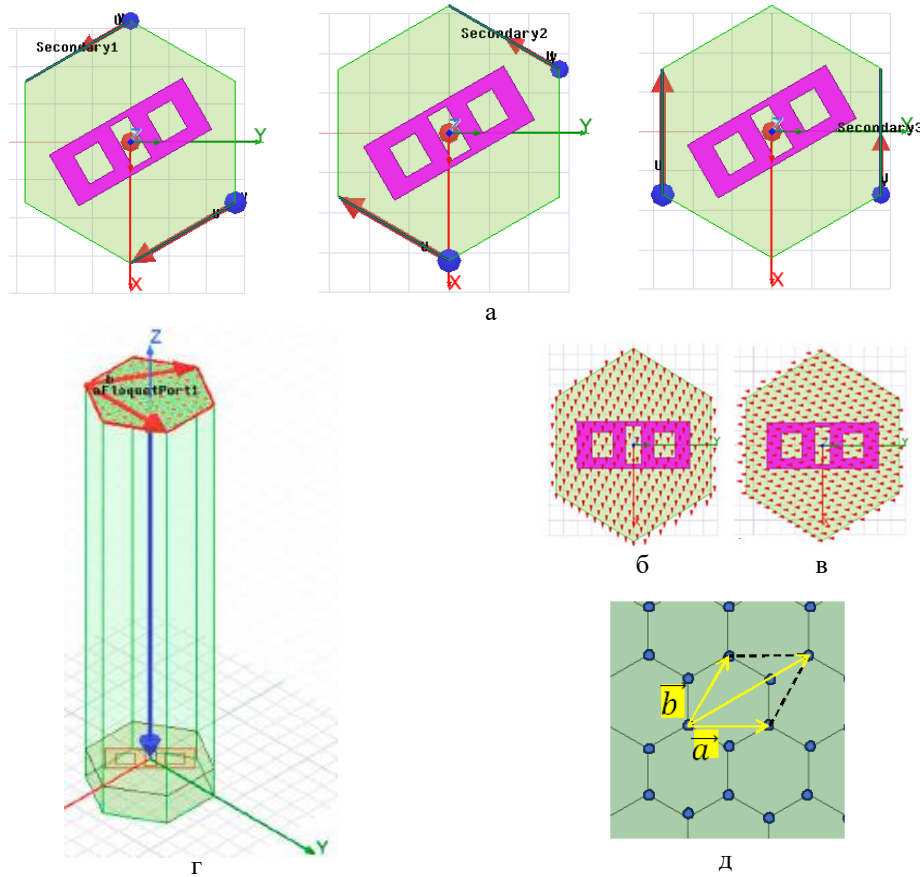


Рис. 2. Постановка периодических граничных условий (а) и порта Флоке (б-д)

В соответствии с теорией бездисперсной ПБ-фазы [23, 24] мета-частицы должны создавать в канале Флоке противофазное отражение собственных мод линейных поляризаций независимо от угла поворота β [11]. Это условие гарантирует линейный закон изменения ПБ-фазы $\psi_{PB} = \pm 2\beta$ для право- и лево-поляризованных волн, соответственно.

Сравним угло-частотные характеристики отражения КП-волн в каналах Флоке ПБ-ячеек гексагональной и квадратной формы. Симуляция показала, что переход от квадратной к более изотропной гексагональной ячейке улучшает угло-частотные зависимости ПБ-фазы от угла поворота β (рис. 3, 4). Собственные моды Флоке противофазны в широкой полосе частот (рис. 3).

Из рис. 4 видно, что вращение мета-частицы в «соте» реализует линейный закон изменения ПБ-фазы от угла β в полосе частот от 9 до 19,5 ГГц. Для квадратной ячейки этот закон отличается от линейного на нижних частотах (9 ГГц, 10 ГГц) и ухудшает поляризационную конверсию КП-волн.

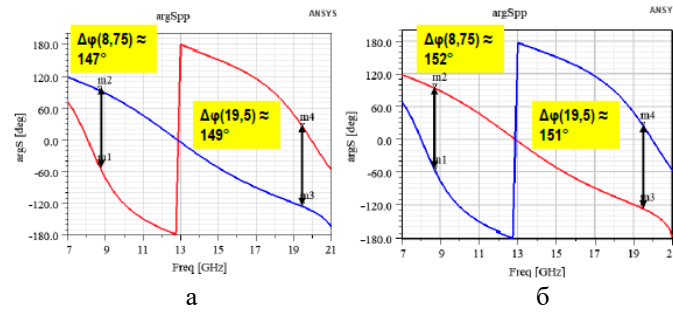


Рис. 3. Фазы отражения собственных мод Флоке линейных X- и Y-поляризаций в полосе от 7 до 21 ГГц при ориентации мета-частиц вдоль осей симметрии ячеек X (a) и Y (б)

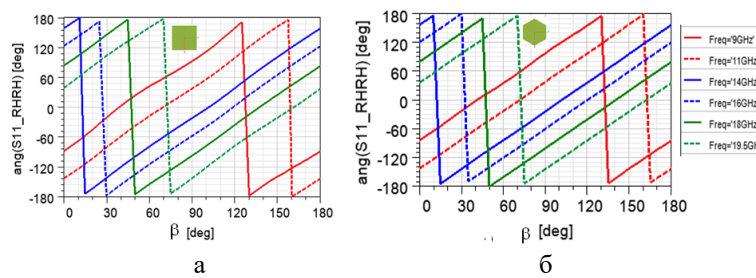


Рис. 4. Зависимости фазы отражения ко-поляризованной правой КП-волны от угла поворота β мета-частиц в ячейках квадратной (a) и гексагональной (б) формы

Частотные характеристики (ЧХ) поляризационной конверсии КП-волн иллюстрирует рис. 5 для разных углов поворота мета-частиц $\beta=0^\circ$; 30° ; 45° вдоль характерных осей симметрии ячеек. «Сота» эффективно отражает ко-поляризованные КП-волны (с потерями менее 0,35 дБ) независимо от угла β в полосе от 8,5 до 19,7 ГГц (рис. 5,б,в). Для квадратной ячейки поляризационная конверсия зависит от β на нижних частотах (рис. 5,а) [14].

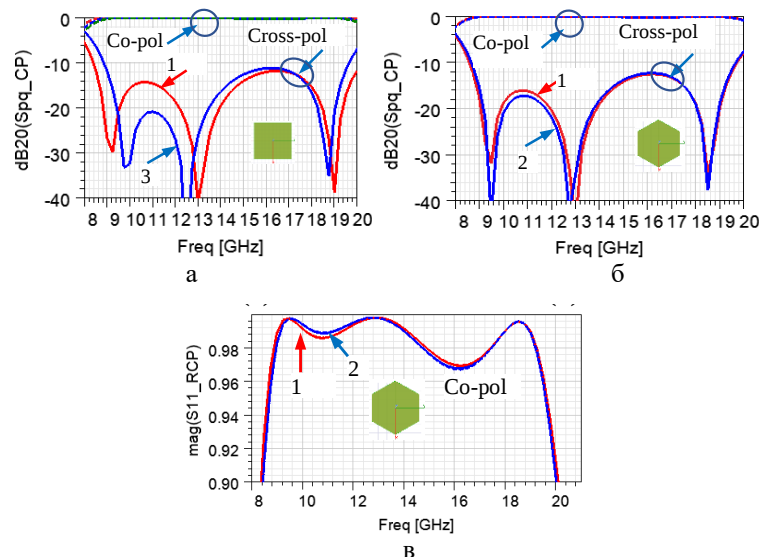


Рис. 5. ЧХ амплитуд коэффициентов отражения КП-волн при углах поворота $\beta=0^\circ$ (линия 1); 30° (линия 2) и 45° (линия 3) в ячейках квадратной (a) и гексагональной (б, в) формы

2. Топология и характеристики рассеяния модулей МП с генерацией ОУМ.

Сравним эффективность фазового гашения обратного рассеяния от ОУМ-модулей из ячеек разной формы за счет генерации вихревых волн. Такие модули из квадратных ячеек имели размер 120×120 мм и описаны нами в [22]. Разработанные ОУМ-модули из «сот» обозначены ниже Mh0p, Mh90p, Mh0n, Mh90n и имели размер 108,8×110,85 мм. Их топология изображена на рис. 6 (границы «сот» показаны лишь на рис. 6,а).

Углы поворота β мета-частиц в ячейках вычислялись по формуле [22]:

$$\beta(x, y) = 0.5 q \cdot \arctan\left(\frac{y}{x}\right) + \beta_0,$$

где q – параметр, определяющий топологический заряд вихря; x, y – координаты центров ячеек в системе координат, связанной с МП, β_0 – начальный угол наклона мета-частиц.

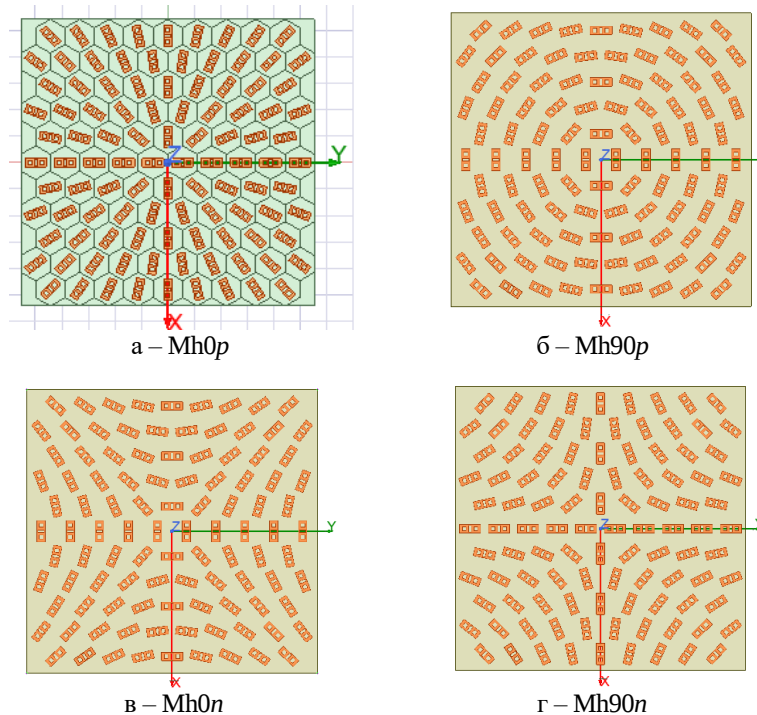


Рис. 6. Топология модулей из гексагональных ячеек с разными фазовыми ОУМ-профилями

Модули Mh0p, Mh90p, Mh0n, Mh90n характеризовались разными параметрами ОУМ $\beta_0 = 0^\circ; 90^\circ$, $q = \pm 2$, которые включены в обозначения модулей (p – positive для $q > 0$ и n – negative для $q < 0$). Модули с квадратными ячейками Ms0p, Ms90p, Ms0n, Ms90n имели аналогичные параметры ОУМ [14].

Симуляция подтвердила, что гексагональные ОУМ-модули эффективнее. Они снижают обратное рассеяние в полосе от 8,5 до 20,8 ГГц более чем на 10 дБ (рис. 7,а,в), а аналогичные модули из квадратных ячеек – лишь на 8 дБ (рис. 7,б,г).

Физическими механизмами фазового гашения поля рассеяния метаповерхностей являются две конверсии. Во-первых, это поляризационная конверсия КП-волн, которая преобразует падающую КП-волну в интенсивную ко-поляризованную и ослабленную кросс-поляризованную компоненты отраженной КП-волны. Эта конверсия определяется, в основном, свойствами спроектированной единичной ячейки с мета-частицей. Во-вторых, это спин-орбитальная конверсия КП-волн при рассеянии от ОУМ-модулей, когда спиновый угловой момент падающей волны круговой поляризации преобразуется в орбитальный угловой момент ко-поляризованного дальнего поля и в спиновый угловой момент кросс-поляризованного дальнего поля [17]. Обратное рассеяние снижается за

счет генерации воронкообразных вихревых волн (рис. 8,а,в) с интенсивной модой ОУМ $k = -1$ (для Mh0p; рис. 8,б) или $k = +3$ (для Mh0n; рис. 8,г). Аналогичные вихри генерируют модули Mh90p и Mh90n. Порядок мод ОУМ совпадает для сравниваемых модулей из гексагональных и квадратных ячеек [14]. При этом вихревое рассеяние сохраняется в сверхширокой полосе частот от 8,5 до 21 ГГц (рис. 9).

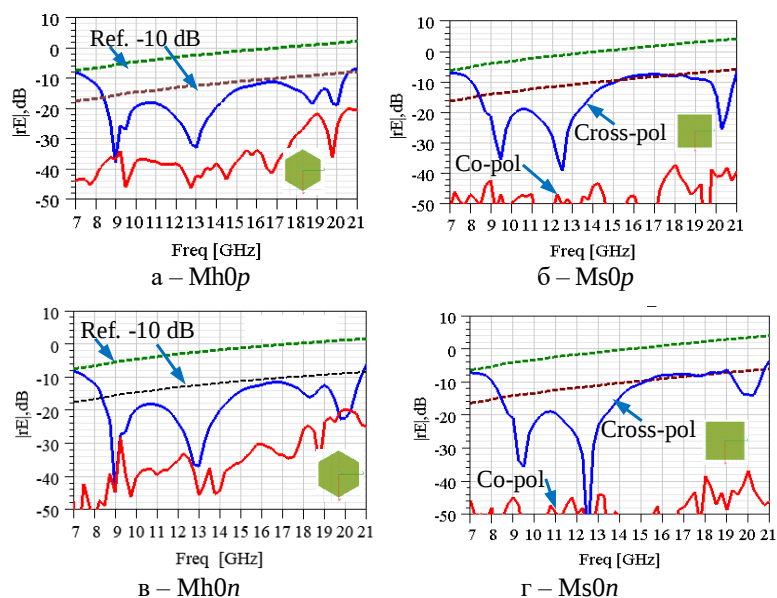


Рис. 7. Гашение обратного рассеяния КР-волн от модулей с гексагональными (а, в) и квадратными (б, г) ячейками; зеленые линии – кросс-поля эталонов, коричневые – уровень минус 10 дБ от полей эталонов

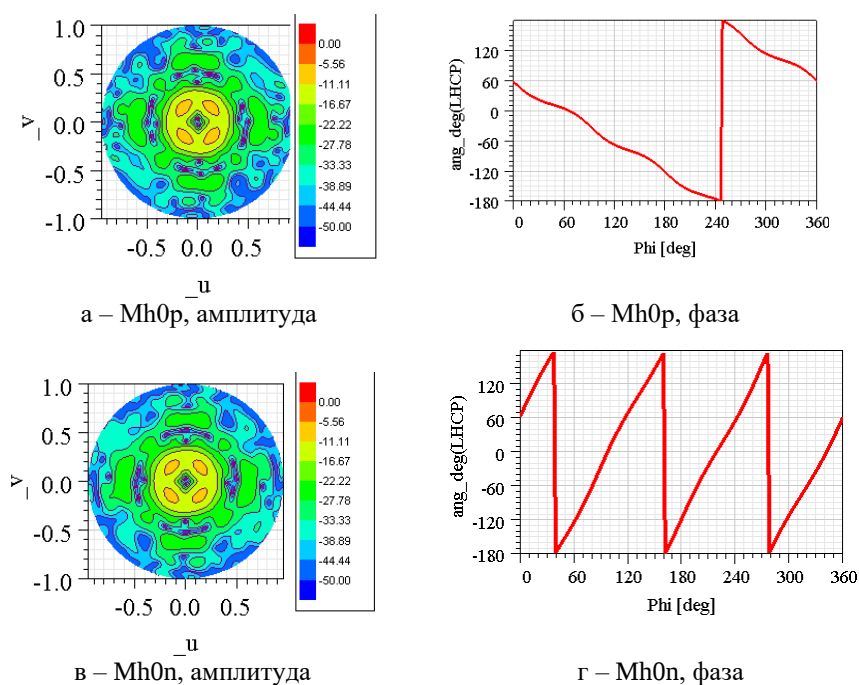


Рис. 8. Амплитудные 2D-диаграммы рассеяния ко-поля модулей (а, в) и распределение фазы рассеянного поля вдоль азимутального угла (б, г) на частоте 14 ГГц

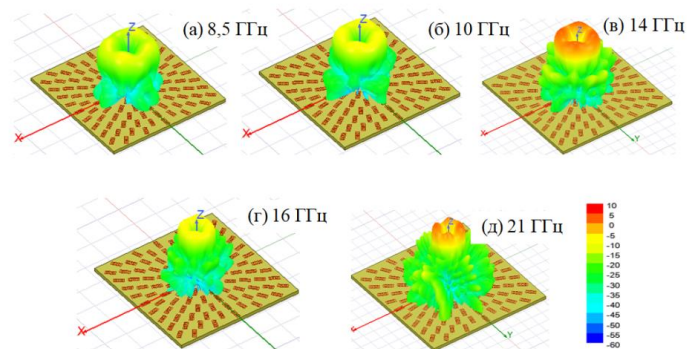


Рис. 9. 3D-диаграммы рассеяния (в дБ) ко-поля модуля $Mh0r$ в верхней полусфере

Заключение. В статье рассмотрена возможность улучшения фазового гашения рассеяния с помощью метаповерхностей Панчаратнама-Берри с генерацией вихревых волн с орбитальным угловым моментом и спиралевидным фазовым фронтом. Метаповерхности выполнены из ячеек гексагональной формы (вместо традиционных квадратных ячеек). Симуляция рассеяния КП-волн подтвердила, что ОУМ-модули МП с гексагональными ячейками эффективнее снижают кросс-поляризованное поле обратного рассеяния (свыше 10 дБ) по сравнению с МП из квадратных ячеек (лишь на 8 дБ) в полосе от 8,5 до 20,8 ГГц. Полученные результаты могут быть полезны при выборе формы единичных ячеек метаповерхностей, предназначенных для широкополосного гашения рассеяния.

Исследование выполнено за счет гранта Российского научного фонда № 22-19-00537-П, <https://rscf.ru/project/22-19-00537/> в Центре коллективного пользования «Прикладная электродинамика и антенные измерения» ЮФУ.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Dai H., Zhao Y., Yu C. A multi-elements chessboard random coded metasurface structure for ultra-wideband radar cross section reduction // IEEE Access. – 2020. – Vol. 8. – P. 56462-56468. – <https://doi.org/10.1109/ACCESS.2020.2977630>.
2. Al-Nuaimi M.K.T., Hong W., He Y. Design of diffusive modified chessboard metasurface // IEEE Antennas and Wireless Propagation Letters. – 2019. – Vol. 18, No. 8. – P. 1621-1625. <https://doi.org/10.1109/LAWP.2019.2925378>.
3. Климов А.В., Семенихин А.И. Модели широкополосных отражательных поляризаторов с использованием метаматериалов // Известия ЮФУ. Технические науки. – 2013. – № 11 (148). – С. 151-156.
4. Chen W. et al. Broadband polarization conversion metasurface for radar cross section reduction // 2018 International Conference on Microwave and Millimeter Wave Technology (ICMMT). – IEEE, 2018. – P. 1-3. – <https://doi.org/10.1109/ICMMT.2018.8563649>.
5. Dong G. et al. Wideband polarization-independent anomalous reflection mediated by metasurface // 2016 11th International Symposium on Antennas, Propagation and EM Theory (ISAPE). – IEEE, 2016. – P. 753-755. – <https://doi.org/10.1109/ISAPE.2016.7834068>.
6. Semikhina A.I., Semikhina D.V. Wideband phase gradient metasurfaces with anomalous diffuse scattering to reduce RCS for linear and circular polarizations // 2021 Radiation and Scattering of Electromagnetic Waves (RSEMW). – IEEE, 2021. – P. 261-264. – <https://doi.org/10.1109/RSEMW52378.2021.9494040>.
7. Chen K. et al. Geometric phase coded metasurface: from polarization dependent directive electromagnetic wave scattering to diffusion-like scattering // Scientific reports. – 2016. – Vol. 6, No. 1. – P. 35968. – <https://doi.org/10.1038/srep35968>.
8. Семенихин А.И., Климов А.В., Савицкий А.Н. Анизотропная импедансная цилиндрическая метаповерхность для самоадаптивного гашения рассеянных волн любых поляризаций // Известия ЮФУ. Технические науки. – 2023. – № 6. – С. 258-267. – <https://doi.org/10.18522/2311-3103-2023-6-258-267>.
9. Chen H.T., Taylor A.J., Yu N. A review of metasurfaces: physics and applications // Reports on progress in physics. – 2016. – Vol. 79, No. 7. – P. 076401. – <https://iopscience.iop.org/article/10.1088/0034-4885/79/7/076401>.

10. Tian Y. *et al.* Manipulation of the arbitrary scattering angle based on all-dielectric transmissive Pancharatnam Berry phase coding metasurfaces in the visible range // *Optics Express*. – 2020. – Vol. 28, No. 21. – P. 32107-32123. – <https://doi.org/10.1364/OE.409509>.
11. Wu X. *et al.* Ultra-broadband Pancharatnam-Berry phase metasurface for arbitrary rotation of linear polarization and beam splitter // *Optics express*. – 2022. – Vol. 30, No. 9. – P. 15158-15171. – <https://doi.org/10.1364/OE.456393>.
12. Li B. Q. *et al.* Electromagnetic Scattering Suppression Based on Multi-beam OAM Metasurface // 2022 International Conference on Microwave and Millimeter Wave Technology (ICMMT). – IEEE, 2022. – P. 1-3. – <https://doi.org/10.1109/ICMMT55580.2022.10023076>.
13. Liu Q. *et al.* RCS reduction metasurface based on orbital angular momentum // *Results in Physics*. – 2023. – Vol. 53. – P. 107008. – <https://doi.org/10.1016/j.rinp.2023.107008>.
14. Semenikhin A.I. *et al.* Wide-Angle Broadband Cancellation of Scattering from Metasurfaces with OAM and Combined Phase Profiles // 2024 International Conference on Electromagnetics in Advanced Applications (ICEAA). – IEEE, 2024. – P. 077-080. – <https://doi.org/10.1109/ICEAA61917.2024.10701700>.
15. Семенихин А.И., Семенихина Д.В. Метаповерхности Патчаратнама-Берри с генерацией углового орбитального момента и комбинированным фазовым кодированием для широкополосного широкоугольного снижения ЭПР // *Журнал радиоэлектроники*. – 2024. – №. 5. – <https://doi.org/10.30898/1684-1719.2024.5.9>.
16. Semenikhin A.I., Semenikhina D.V., Yukhanov Y.V. Digital Pancharatnam-Berry metasurfaces with 1-bit OAM-modules for broadband RCS reduction // 2023 International Conference on Electromagnetics in Advanced Applications (ICEAA). – IEEE, 2023. – P. 019-023. – <https://doi.org/10.1109/ICEAA57318.2023.10297857>.
17. Yu H., Su J., Li Z. High-gain lens antenna using transmissive phase gradient metasurface // 2018 IEEE 4th International Conference on Computer and Communications (ICCC). – IEEE, 2018. – P. 1032-1036. – <https://doi.org/10.1109/CompComm.2018.8780901>.
18. Sun G. C. *et al.* Exciting Extended Bound States in the Continuum in Symmetry-Broken Scalable All Dielectric THz Metasurface // 2023 Cross Strait Radio Science and Wireless Technology Conference (CSRSWTC). – IEEE, 2023. – P. 1-2. – <https://doi.org/10.1109/CSRSWTC60855.2023.10426942>.
19. Yang Y., Zhang W. Low Sidelobe Level Transmitarray Antenna Using Hexagonal Unit Cell // 2023 International Applied Computational Electromagnetics Society Symposium (ACES-China). – IEEE, 2023. – P. 1-3. – <https://doi.org/10.23919/ACES-China60289.2023.10250043>.
20. Zou Q. *et al.* An Orthogonal Circular Polarization Converting Transmitarray Using Hexagonal Arrangement // 2024 International Conference on Microwave and Millimeter Wave Technology (ICMMT). – IEEE, 2024. – Vol. 1. – P. 1-3. – <https://doi.org/10.1109/ICMMT61774.2024.10671830>.
21. Zhang W. *et al.* Broadband metasurface antenna using hexagonal loop-shaped unit cells // *IEEE access*. – 2020. – Vol. 8. – P. 223797-223805. – <https://doi.org/10.1109/ACCESS.2020.3043656>.
22. Du J. *et al.* Optical vortex array: generation and applications // *Chinese Optics Letters*. – 2024. – Vol. 22, No. 2. – P. 020011. – <https://opg.optica.org/col/abstract.cfm?URI=col-22-2-020011>.
23. Pancharatnam S. Generalized theory of interference and its applications // *Proceedings of the Indian Academy of Sciences-section A*. – New Delhi: Springer India, 1956. – Vol. 44, No. 6. – P. 398-417. – <https://link.springer.com/article/10.1007/BF03046095>.
24. Berry M.V. Quantal phase factors accompanying adiabatic changes // *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*. – 1984. – Vol. 392, No. 1802. – P. 45-57. – <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1984.0023>.

REFERENCES

1. Dai H., Zhao Y., Yu C. A multi-elements chessboard random coded metasurface structure for ultra-wideband radar cross section reduction, *IEEE Access*, 2020, Vol. 8, pp. 56462-56468. Available at: <https://doi.org/10.1109/ACCESS.2020.2977630>.
2. Al-Nuaimi M.K.T., Hong W., He Y. Design of diffusive modified chessboard metasurface, *IEEE Antennas and Wireless Propagation Letters*, 2019, Vol. 18, No. 8, pp. 1621-1625. Available at: <https://doi.org/10.1109/LAWP.2019.2925378>.
3. Klimov A.V., Semenikhin A.I. Modeli širokopolosnykh otrazhatel'nykh polarizatorov s ispol'zovaniem metamaterialov [Models of broadband reflective polarizers using metamaterials], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2013, No. 11 (148), pp. 151-156.
4. Chen W. *et al.* Broadband polarization conversion metasurface for radar cross section reduction, 2018 International Conference on Microwave and Millimeter Wave Technology (ICMMT). IEEE, 2018, pp. 1-3. Available at: <https://doi.org/10.1109/ICMMT.2018.8563649>.

5. Dong G. et al. Wideband polarization-independent anomalous reflection mediated by metasurface, *2016 11th International Symposium on Antennas, Propagation and EM Theory (ISAPE)*. IEEE, 2016, pp. 753-755. Available at: <https://doi.org/10.1109/ISAPE.2016.7834068>.
6. Semenikhin A.I., Semenikhina D.V. Wideband phase gradient metasurfaces with anomalous diffuse scattering to reduce RCS for linear and circular polarizations, *2021 Radiation and Scattering of Electromagnetic Waves (RSEMW)*. IEEE, 2021, pp. 261-264. Available at: <https://doi.org/10.1109/RSEMW52378.2021.9494040>.
7. Chen K. et al. Geometric phase coded metasurface: from polarization dependent directive electromagnetic wave scattering to diffusion-like scattering, *Scientific reports*, 2016, Vol. 6, No. 1, pp. 35968. Available at: <https://doi.org/10.1038/srep35968>.
8. Semenikhin A.I., Klimov A.V., Savitskiy A.N. Anizotropnaya impedansnaya tsilindricheskaya metapoverkhnost' dlya samoadaptivnogo gasheniya rasseyannykh voln lyubykh polarizatsiy [Anisotropic impedance cylindrical meta-surface for self-adaptive suppression of scattered waves of any polarizations], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2023, No. 6, pp. 258-267. Available at: <https://doi.org/10.18522/2311-3103-2023-6-258-267>.
9. Chen H.T., Taylor A.J., Yu N. A review of metasurfaces: physics and applications, *Reports on progress in physics*, 2016, Vol. 79, No. 7, pp. 076401. Available at: <https://iopscience.iop.org/article/10.1088/0034-4885/79/7/076401>.
10. Tian Y. et al. Manipulation of the arbitrary scattering angle based on all-dielectric transmissive Pancharatnam Berry phase coding metasurfaces in the visible range, *Optics Express*, 2020, Vol. 28, No. 21, pp. 32107-32123. – <https://doi.org/10.1364/OE.409509>.
11. Wu X. et al. Ultra-broadband Pancharatnam-Berry phase metasurface for arbitrary rotation of linear polarization and beam splitter, *Optics express*, 2022, Vol. 30, No. 9, pp. 15158-15171. Available at: <https://doi.org/10.1364/OE.456393>.
12. Li B. Q. et al. Electromagnetic Scattering Suppression Based on Multi-beam OAM Metasurface, *2022 International Conference on Microwave and Millimeter Wave Technology (ICMMT)*. IEEE, 2022, pp. 1-3. Available at: <https://doi.org/10.1109/ICMMT55580.2022.10023076>.
13. Liu Q. et al. RCS reduction metasurface based on orbital angular momentum, *Results in Physics*, 2023, Vol. 53, pp. 107008. Available at: <https://doi.org/10.1016/j.rinp.2023.107008>.
14. Semenikhin A.I. et al. Wide-Angle Broadband Cancellation of Scattering from Metasurfaces with OAM and Combined Phase Profiles, *2024 International Conference on Electromagnetics in Advanced Applications (ICEAA)*. IEEE, 2024, pp. 077-080. Available at: <https://doi.org/10.1109/ICEAA61917.2024.10701700>.
15. Semenikhin A.I., Semenikhina D.V. Metapoverkhnosti Patcharatnama-Berri s generatsiey uglovogo orbital'nogo momenta i kombinirovannym fazovym kodirovaniem dlya shirokopolosnogo shirokougol'nogo snizheniya EPR [Patcharatnam-Berry metasurfaces with angular orbital momentum generation and combined phase encoding for broadband wide-angle RCS reduction], *Zhurnal radioelektroniki* [Journal of Radio Electronics], 2024, No. 5. Available at: <https://doi.org/10.30898/1684-1719.2024.5.9>.
16. Semenikhin A.I., Semenikhina D.V., Yukhanov Y.V. Digital Pancharatnam-Berry metasurfaces with 1-bit OAM-modules for broadband RCS reduction, *2023 International Conference on Electromagnetics in Advanced Applications (ICEAA)*. IEEE, 2023, pp. 019-023. Available at: <https://doi.org/10.1109/ICEAA57318.2023.10297857>.
17. Yu H., Su J., Li Z. High-gain lens antenna using transmissive phase gradient metasurface, *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*. IEEE, 2018, pp. 1032-1036. Available at: <https://doi.org/10.1109/CompComm.2018.8780901>.
18. Sun G. C. et al. Exciting Extended Bound States in the Continuum in Symmetry-Broken Scalable All Dielectric THz Metasurface, *2023 Cross Strait Radio Science and Wireless Technology Conference (CSRSWTC)*. IEEE, 2023, pp. 1-2. Available at: <https://doi.org/10.1109/CSRSWTC60855.2023.10426942>.
19. Yang Y., Zhang W. Low Sidelobe Level Transmitarray Antenna Using Hexagonal Unit Cell, *2023 International Applied Computational Electromagnetics Society Symposium (ACES-China)*. IEEE, 2023, pp. 1-3. Available at: <https://doi.org/10.23919/ACES-China60289.2023.10250043>.
20. Zou Q. et al. An Orthogonal Circular Polarization Converting Transmitarray Using Hexagonal Arrangement, *2024 International Conference on Microwave and Millimeter Wave Technology (ICMMT)*. IEEE, 2024, Vol. 1, pp. 1-3. Available at: <https://doi.org/10.1109/ICMMT61774.2024.10671830>.
21. Zhang W. et al. Broadband metasurface antenna using hexagonal loop-shaped unit cells, *IEEE access*, 2020, Vol. 8, pp. 223797-223805. Available at: <https://doi.org/10.1109/ACCESS.2020.3043656>.
22. Du J. et al. Optical vortex array: generation and applications, *Chinese Optics Letters*, 2024, Vol. 22, No. 2, pp. 020011. Available at: <https://opg.optica.org/col/abstract.cfm?URI=col-22-2-020011>.

23. Pancharatnam S. Generalized theory of interference and its applications, *Proceedings of the Indian Academy of Sciences-section A*. New Delhi: Springer India, 1956, Vol. 44, No. 6, pp. 398-417. Available at: <https://link.springer.com/article/10.1007/BF03046095>.
24. Berry M.V. Quantal phase factors accompanying adiabatic changes, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 1984, Vol. 392, No. 1802, pp. 45-57. Available at: <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1984.0023>.

Семенihin Андрей Илларионович – Южный федеральный университет; e-mail: anilsem@mail.ru; г. Таганрог, Россия; тел.: 89061831649; кафедра антенн и радиопередающих устройств; доцент; SPIN: 5609-3793.

Семенихина Диана Викторовна – Южный федеральный университет; e-mail: d_semenikhina@mail.ru; г. Таганрог, Россия; тел.: 89094412058; кафедра антенн и радиопередающих устройств; профессор; SPIN: 8694-5179.

Зыкина Анна Михайловна – Южный федеральный университет; e-mail: a.zikina@yandex.ru; г. Таганрог, Россия; тел.: 89615866362; магистрант.

Semenikhin Andrey Illarionovich – Southern Federal University; e-mail: anilsem@mail.ru; Taganrog, Russia; phone: +79061831649; the department of A&RTD; associate professor; SPIN: 5609-3793.

Semenikhina Diana Viktorovna – Southern Federal University; e-mail: d_semenikhina@mail.ru; Taganrog, Russia; phone: +79094412058; the department of A&RTD; professor; SPIN: 8694-5179.

Zikina Anna Mihailovna – Southern Federal University; e-mail: a.zikina@yandex.ru; Taganrog, Russia; phone: +79615866362; magister.

УДК 621.396.677.3

DOI 10.18522/2311-3103-2026-1-218-224

В.О. Игнатович, Н.Н. Кисель

МОДЕЛИРОВАНИЕ И ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК МИКРОПОЛОСКОВОЙ МИМО-АНТЕННОЙ СИСТЕМЫ

Представлены результаты комплексного исследования характеристик микрополосковой МИМО-антенной системы, предназначенной для использования в телекоммуникационных средствах пятого поколения в диапазоне частот $n79$ (4,4–5,0 ГГц). Целью исследования являлась разработка компактной антенной решётки конфигурации 1×2 на единой диэлектрической подложке и оценка её электродинамических свойств с учётом факторов, способных оказывать влияние на параметры согласования и излучения в реальных условиях эксплуатации. Конструкция антенны включает два излучателя, четыре порта возбуждения и модифицированный слой заземления, обеспечивающий расширенную функциональность и устойчивость рабочих характеристик. В ходе численного моделирования установлено, что предложенная антенная система обеспечивает стабильные значения коэффициента стоячей волны, высокий коэффициент усиления и согласованные диаграммы направленности во всём рабочем диапазоне частот. Особое внимание уделено анализу влияния внешних воздействий, в частности попадания тонкой водяной плёнки на поверхность излучателя. Показано, что присутствие влаги приводит к росту КСВ более чем на 0,5, снижению КПД и ухудшению направленных свойств антенны на верхней границе диапазона, что подчёркивает необходимость использования защитных конструктивных решений при эксплуатации изделия. Дополнительно исследована корректность замены меди идеальным проводником при моделировании. Разница в значениях КСВ не превышает 0,006, что подтверждает возможность применения такого упрощения на этапах предварительного проектирования без существенной потери точности. Экспериментальный образец антенны был изготовлен и испытан в лабораторных условиях. Полученные экспериментальные данные продемонстрировали высокую степень совпадения с результатами численного моделирования, что подтверждает валидность предложенной математической модели и практическую применимость разработанной МИМО-антенной системы в инфраструктуре сетей связи нового поколения.

МИМО антенны; микрополосковые антенны; мобильная связь; 5G; антенные решетки; телекоммуникационные системы.

V.O. Ignatovich, N.N. Kisel

**MODELLING AND EXPERIMENTAL RESEARCH OF MICROSTRIP
MIMO-ANTENNA SYSTEM'S CHARACTERISTICS**

The paper presents the results of a comprehensive study of the characteristics of a microstrip MIMO antenna system intended for application in fifth-generation communication equipment operating within the n79 frequency band (4.4–5.0 GHz). The purpose of the research was to develop a compact 1×2 antenna array on a single dielectric substrate and to evaluate its electrodynamic properties, taking into account factors capable of influencing matching and radiation parameters under real operating conditions. The proposed design incorporates two radiating elements, four excitation ports, and a modified ground plane, which provides enhanced functional stability and improved operational characteristics. Numerical modelling demonstrated that the developed antenna system ensures stable standing-wave ratio values, high antenna gain, and consistent radiation patterns across the entire operating frequency band. Particular attention was devoted to assessing the influence of external environmental effects, especially the formation of a thin water film on the radiating surface. The study revealed that the presence of moisture results in an SWR increase of more than 0.5, a decrease in antenna efficiency, and degradation of radiation characteristics at higher frequencies, highlighting the necessity of employing protective structural solutions when operating under humidity-prone conditions. Additionally, the feasibility of replacing copper with an ideal conductor in electromagnetic simulations was examined. The difference in SWR between the realistic and idealized models did not exceed 0.006, confirming that such a simplification may be applied during early design stages without significant loss of accuracy. An experimental prototype of the antenna was fabricated and tested under laboratory conditions. The obtained measurements demonstrated a high degree of agreement with the modelling results, validating the proposed mathematical model and confirming the practical applicability of the developed MIMO antenna system in next-generation communication infrastructures.

MIMO antennas; microstrip patch antenna; network; 5G; antenna arrays; telecommunication systems.

Введение. В настоящее время активно развиваются системы связи 5G, использующие частотный диапазон до 6 ГГц (макросоты), которые позволяют реализовать показатели систем связи на порядок лучшие по сравнению со стандартом 4G. Микрополосковые антенны являются наиболее подходящими антеннами для мобильных объектов, поскольку имеют ряд преимуществ, такие как компактность, технологичность изготовления, а использование технологии МИМО позволяет увеличить скорость передачи данных и качество сигнала за счет большего количества антенн. Наиболее удобным диапазоном для 5G связи следует отнести диапазон n79: 4,4–5,0 ГГц. Требования к антеннам, диапазонам частот и мощности излучения меняется, что связано с многоэтапным процессом внедрения стандарта 5G.

Исследование влияния внешних воздействий на характеристики МИМО-антенны. В [1] приведены результаты моделирования основных характеристик МИМО-антенной решетки 1×2 в диапазоне 4,4–5,0 ГГц и исследованы чувствительности характеристик антенны к изменениям ее геометрических и электрофизических параметров подложки: Rogers RO3006 [2], Arlon TC600 [3], FSD615T (углеводородный материал с нанокерамикой и стекловолокном) [4], проанализировано влияние фазовых соотношений в системе питания на диаграммы направленности антенной решетки. В литературе [5–24] так же рассматривались вопросы применения различных видов антенных систем, в том числе, микрополосковых, в современных системах мобильной связи. В рассмотренных материалах так же уделялось внимание технологии МИМО, отмечался высокий потенциал данной технологии в современных системах беспроводной связи.

В продолжении исследований, проведенных в [1], было проанализировано влияние на направленные и частотные свойства антенны попадания влаги на поверхности излучателя. Для этого рассматривался однородный слой водяной пленки толщиной 0.035 мм на поверхности излучателя.

График зависимости коэффициента отражения от частоты и диаграмма направленности антенны при наличии и отсутствии воды на поверхности представлены на рис. 1.

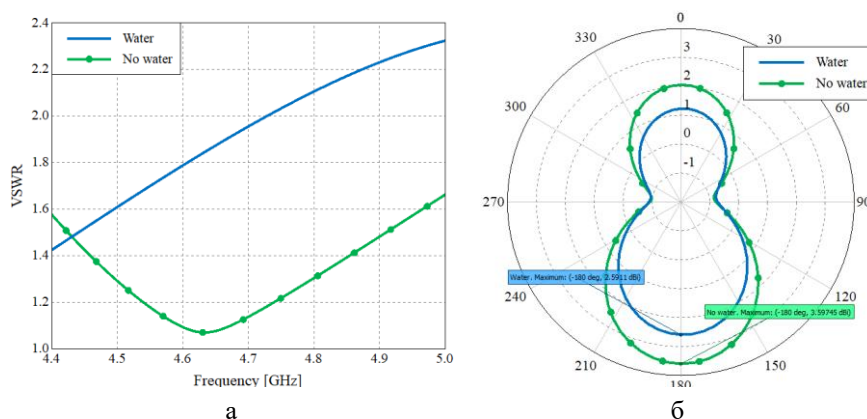


Рис. 1. Частотная зависимость КСВ (а) и диаграмма направленности на фиксированной частоте при наличии и отсутствии воды на поверхности излучателя

КСВ антенны увеличивается, с ростом частоты наблюдается практически линейный его рост с 1,4 для 4,4 ГГц до 2,3 на 5,0 ГГц. В верхней части спектра значение КПД антенны при наличии водяной пленки становится меньше 0,7 и наблюдается снижение коэффициента усиления.

Экспериментальное исследование ММО-антенной системы. На основе модели [1] был создан образец ММО-антенны с четырьмя SMA-разъемами (рис. 2). Габаритные размеры антенны 27 мм x 27 мм.

Экспериментальные исследования были проведены в Центре инженерных разработок «Микроволновые технологии электродинамических структур» Южного федерального университета. В рамках исследования ММО-антенной системы были проведены измерения КСВ антенны в зависимости от частоты. Измерения проводились с помощью векторного анализатора цепей. Фотография изготовленной антенны с подключенными SMA-разъемами (вид сверху) представлена на рис. 2.

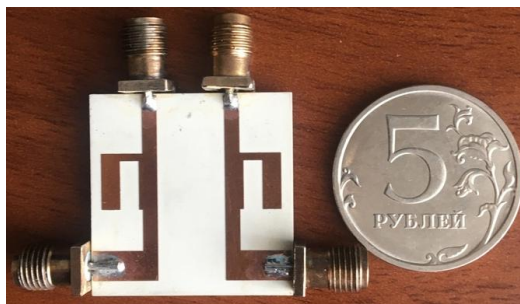


Рис. 2. Фотография изготовленной антенны с подключенными SMA-разъемами (вид сверху)

Так как антенна имеет два порта питания, для измерений задействованы оба порта генератора. Режимы измерений – S_{11} для порта №1, S_{22} для порта №2. Результаты измерений коэффициента отражения в сравнении с результатами моделирования представлены на рис. 3.

Экспериментальная и расчетная характеристика коэффициента отражения полностью согласованы, минимальные значения коэффициента отражения наблюдаются на частоте 4,65 ГГц. Во всем диапазоне частот 4,4–5,0 ГГц коэффициент отражения ниже 12 дБ. Наблюдаемые отличия можно объяснить возможным разбросом параметров диэлектрической подложки, отсутствием учета частотной дисперсии электрофизических параметров подложки, наличием паразитных емкостей в области SMA-разъемов.

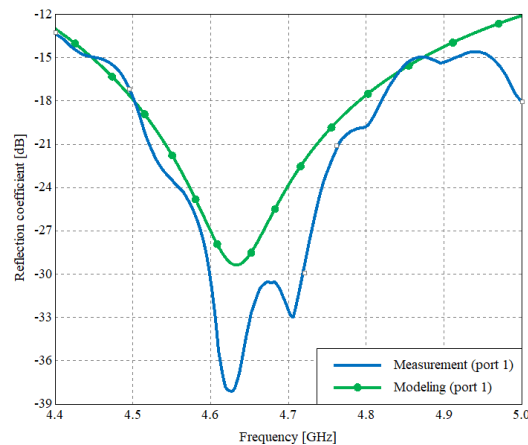


Рис. 3. График зависимости коэффициента отражения ММО-антенны от частоты (результаты моделирования и результаты измерения параметров изготовленного образца)

Графики зависимости КСВ ММО-антенны от частоты, полученные при измерении и при компьютерном моделировании, представлены на рис. 4.

Результаты количественно хорошо согласуются друг с другом, что можно рассматривать как верификацию результатов, полученных с использованием САПР Altair FEKO и как корректность математической модели антенны.

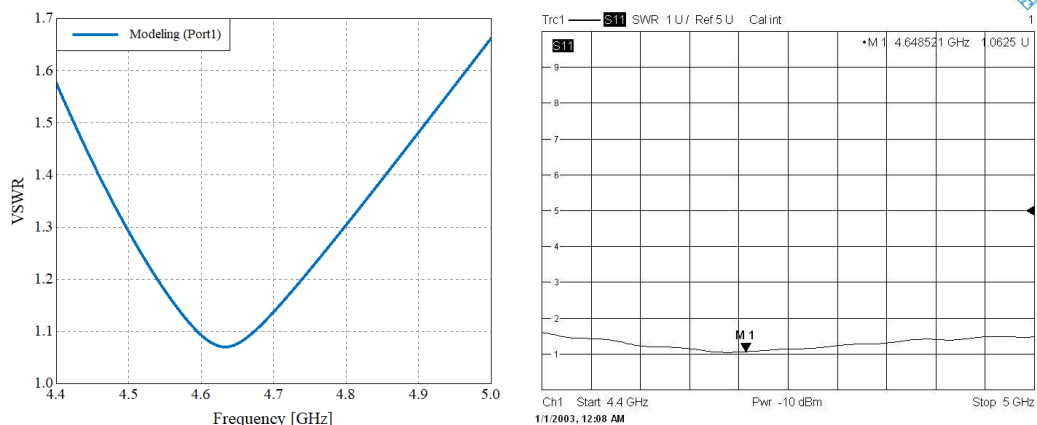


Рис. 4. Графики зависимости КСВ ММО-антенны от частоты, полученные в ходе моделирования и в ходе измерения

На рис. 5 приведены результаты исследования диаграммы направленности ММО-антенны. Измерение проводилось в безэховой камере.

Диаграммы направленности в плоскости xOz , полученные в ходе моделирования и в процессе измерения, представлены на рис. 5. Обе диаграммы были получены при равно-амплитудном и синфазном возбуждении. Антенна была размещена на поворотном устройстве коннекторами питания вниз.

Как видно из приведенных результатов, значения максимумов излучения и их направления приблизительно совпадают. Имеются те же два максимума излучения, уровень сигнала в которых отличается менее чем на 1 dBi. В экспериментальной диаграмме направленности наблюдаются небольшие осцилляции, что, по-видимому, можно объяснить наличием нескольких SMA-разъемов и кабелей от них, которые при угловом повороте антенны меняются свое положение, которое затруднительно контролировать в ходе эксперимента.

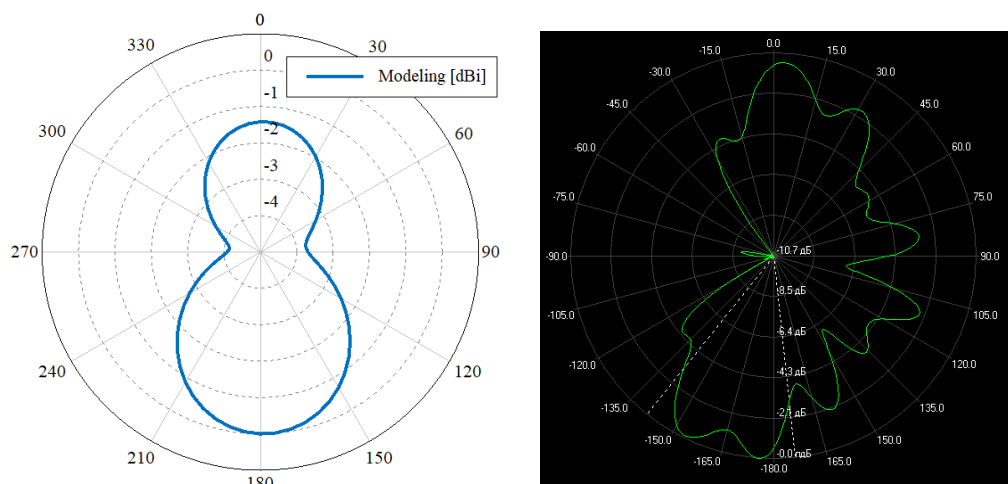


Рис. 5. Диаграммы направленности в плоскости xOz , полученные в ходе моделирования и в ходе измерения (нормированные к максимуму)

Заключение. Таким образом, в работе приведены результаты анализа влияния водяной пленки на поверхности излучателя. Моделирование показало, что КСВ антенны возрастает более чем на 0.5, по сравнению с результатами модели без водяной пленки. Одновременно с этим снижается КПД антенны, ухудшаются ее направленные и частотные свойства. В связи с этим требуется обязательное использование обтекателя для снижения воздействия внешних факторов на характеристики антенны.

Экспериментальное исследование разработанного реального образца антенны показало хорошее согласование полученных характеристик с результатами численного моделирования. Сравнивались коэффициент отражения, КСВ и диаграмма направленности антенной решетки. Измерения проводились в диапазоне п79, который соответствует принятому в РФ диапазону для 5G.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Игнатович В.О., Кисель Н.Н. Моделирование характеристик микрополосковых антенн при применении в технологиях MIMO и beamforming // Матер. Всероссийской научно-технической конференции с международным участием имени профессора О.Н. Пьявченко (КомТех-2024). Таганрог, 5–7 июня 2024. – Т. 1. – С. 220-227. – ISBN 978-5-9275-4718-0.
2. RO3000® Series Circuit Materials. Rogers Corporation. – URL: https://rezonit.ru/upload/spetsifikatsii/Rogers_RO3000.pdf (дата обращения: 27.04.2024).
3. Arlon TC600: Enhanced Thermal Conductivity Ceramic Filled PTFE/Woven Fiberglass Laminate for Microwave Printed Circuit Boards. Arlon LLC. – URL: https://rezonit.ru/upload/spetsifikatsii/Arlon_TC600.pdf (дата обращения: 27.07.2024).
4. FSD615T – прямой аналог материала RO4360G2. ООО «Элтех-М». – URL: <https://elmt.ru/svch-materialy-fsd.html/nid/2099> (дата обращения: 15.03.2024).
5. Stallings W. 5G Wireless: A Comprehensive Introduction. – Boston: Addison-Wesley, 2021. – 672 p.
6. Тонг Вэнь, Чжу Пейин. Сети 6G: путь от 5G к 6G глазами разработчиков. – М.: ДМК Пресс, 2022. – 624 p. – ISBN: 978-5-97060-995-8.
7. Cui T.J., Qi M.Q., Wan X., Zhao J., Cheng Q. Coding metamaterials, digital metamaterials and programmable metamaterials // Light: Science & Applications. – 2014. – Vol. 3, No. 10. – e218.
8. NTT DoCoMo. DOCOMO conducts world's first successful trial of transparent dynamic metasurface. – Tokyo, 2020. – URL: <https://www.nttdocomo.co.jp> (дата обращения: 20.04.2024).
9. Suzuki D., Oda S., Kawano Y. A flexible and wearable terahertz scanner // Nature Photonics. – 2016. – Vol. 10, No. 12. – P. 809-813.
10. Suzuki D., Ochiai Y., Kawano Y. Thermal device design for a carbon nanotube terahertz camera // ACS Omega. – 2018. – Vol. 3, No. 3. – P. 3540-3547.
11. Zhang M., Yeow J.T. A flexible, scalable, and self-powered midinfrared detector based on transparent PEDOT:PSS/graphene composite // Carbon. – 2020. – Vol. 156. – P. 339-345.

12. Grzyb J., Heinemann B., Pfeiffer U.R. Solid-state terahertz super-resolution imaging device in 130-nm SiGe BiCMOS technology // IEEE Transactions on Microwave Theory and Techniques. – 2017. – Vol. 65, No. 11. – P. 4357-4372.
13. Grzyb J., Heinemann B., Pfeiffer U.R. A 0.55 THz near-field sensor with μm -range lateral resolution fully integrated in 130 nm SiGe BiCMOS // IEEE Journal of Solid-State Circuits. – 2016. – Vol. 51, No. 12. – P. 3063-3077.
14. Hillger P., Jain R., Grzyb J. et al. A 128-pixel system-on-a-chip for real-time super-resolution terahertz near-field imaging // IEEE Journal of Solid-State Circuits. – 2018. – Vol. 53, No. 12. – P. 3599-3612.
15. Grzyb J., Pfeiffer U.R. THz direct detector and heterodyne receiver arrays in silicon nanoscale technologies // Journal of Infrared, Millimeter, and Terahertz Waves. – 2015. – Vol. 36, No. 10. – P. 998-1032.
16. Filipovic D.F., Gearhart S.S., Rebeiz G.M. Double-slot antennas on extended hemispherical and elliptical silicon dielectric lenses // IEEE Transactions on Microwave Theory and Techniques. – 1993. – Vol. 41, No. 10. – P. 1738-1749.
17. Park S.-G., Choi Y., Oh Y.-J., Jeong K.-H. Terahertz photoconductive antenna with metal nanoislands // Optics Express. – 2012. – Vol. 20, No. 23. – P. 25530-25535.
18. Tanoto H., Teng J., Wu Q., et al. Nano-antenna in a photoconductive photomixer for highly efficient continuous-wave terahertz emission // Scientific Reports. – 2013. – Vol. 3. – P. 2824.
19. Polat E.O., Mercier G., Nikitskiy I., et al. Flexible graphene photodetectors for wearable fitness monitoring // Science Advances. – 2019. – Vol. 5, No. 9.
20. Balanis C.A. Antenna Theory: Analysis and Design. – 4th ed. – Hoboken: John Wiley & Sons, 2016. – 1112 p.
21. Антупов С.А., Ищенко Е.А., Кострова В.Н. и др. Исследование влияния расстояния между антеннами в ММО-антенной решетке для сетей пятого поколения // Вестник Воронежского государственного технического университета. – 2020. – Т. 16, № 4. – С. 91-96.
22. Сазонов Д.М. Антенны и устройства СВЧ. – М.: Высшая школа, 1988. – 432 с.
23. Tian R., Lau B.K., Ying Z. Multiplexing efficiency of MIMO antennas // IEEE Antennas and Wireless Propagation Letters. – 2011. – Vol. 10. – P. 183-186.
24. Lamri I.E., Mansoul A., Nakmouche M.F., Belattar M. Design of Novel UWB 4-element MIMO Microstrip Patch Antenna for Sub-6 GHz 5G Applications // International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET). – 2021. – P. 7-12.

REFERENCES

1. Ignatovich V.O., Kisel' N.N. Modelirovaniye kharakteristik mikropoloskovykh anten pri primenении v tekhnologiyakh MIMO i beamforming [Modelling of microstrip antenna characteristics for MIMO and beamforming technologies], *Mater. Vserossiyskoy nauchno-tekhnicheskoy konferentsii s mezhdunarodnym uchastiem imeni professora O.N. Pyavchenko (KomTekh-2024)* [Proceedings of the Russian Scientific-Technological Conference named after Prof. O.N. Pyavchenko (ComTech-2024)]. Taganrog, June 5–7, 2024, Vol. 1, pp 220-227. ISBN 978-5-9275-4718-0.
2. RO3000® Series Circuit Materials. Rogers Corporation. Available at: https://rezonit.ru/upload/spetsifikatsii/Rogers_RO3000.pdf (accessed 27 April 2024).
3. Arlon TC600: Enhanced Thermal Conductivity Ceramic Filled PTFE/Woven Fiberglass Laminate for Microwave Printed Circuit Boards. Arlon LLC. Available at: https://rezonit.ru/upload/spetsifikatsii/Arlon_TC600.pdf (accessed 27 July 2024).
4. FSD615T – pryamoy analog materiala RO4360G2. ООО «Eltekh-M» [FSD615T – direct analogue of RO4360G2. ELTM Co]. Available at: <https://eltm.ru/svch-materialy-fsd.html/nid/2099> (accessed 15 March 2024).
5. Stallings W. 5G Wireless: A Comprehensive Introduction. Boston: Addison-Wesley, 2021, 672 p.
6. Tong Ven', Chzhu Peyin. Seti 6G: put' ot 5G k 6G glazami razrabotchikov [6G Networks: From 5G to 6G Through Developer's Perspective]. Moscow: DMK Press, 2022, 624 p. ISBN: 978-5-97060-995-8.
7. Cui T.J., Qi M.Q., Wan X., Zhao J., Cheng Q. Coding metamaterials, digital metamaterials and programmable metamaterials, *Light: Science & Applications*, 2014, Vol. 3, No. 10, e218.
8. NTT DoCoMo. DOCOMO conducts world's first successful trial of transparent dynamic metasurface. Tokyo, 2020. Available at: <https://www.nttdocomo.co.jp> (accessed 20 April 2024).
9. Suzuki D., Oda S., Kawano Y. A flexible and wearable terahertz scanner, *Nature Photonics*, 2016, Vol. 10, No. 12, pp. 809-813.
10. Suzuki D., Ochiai Y., Kawano Y. Thermal device design for a carbon nanotube terahertz camera, *ACS Omega*, 2018, Vol. 3, No. 3, pp. 3540-3547.

11. Zhang M., Yeow J.T. A flexible, scalable, and self-powered midinfrared detector based on transparent PEDOT:PSS/graphene composite, *Carbon*, 2020, Vol. 156, pp. 339-345.
12. Grzyb J., Heinemann B., Pfeiffer U.R. Solid-state terahertz super-resolution imaging device in 130-nm SiGe BiCMOS technology, *IEEE Transactions on Microwave Theory and Techniques*, 2017, Vol. 65, No. 11, pp. 4357-4372.
13. Grzyb J., Heinemann B., Pfeiffer U.R. A 0.55 THz near-field sensor with μm -range lateral resolution fully integrated in 130 nm SiGe BiCMOS, *IEEE Journal of Solid-State Circuits*, 2016, Vol. 51, No. 12, pp. 3063-3077.
14. Hillger P., Jain R., Grzyb J. et al. A 128-pixel system-on-a-chip for real-time super-resolution terahertz near-field imaging, *IEEE Journal of Solid-State Circuits*, 2018, Vol. 53, No. 12, pp. 3599-3612.
15. Grzyb J., Pfeiffer U.R. THz direct detector and heterodyne receiver arrays in silicon nanoscale technologies, *Journal of Infrared, Millimeter, and Terahertz Waves*, 2015, Vol. 36, No. 10, pp. 998-1032.
16. Filipovic D.F., Gearhart S.S., Rebeiz G.M. Double-slot antennas on extended hemispherical and elliptical silicon dielectric lenses, *IEEE Transactions on Microwave Theory and Techniques*, 1993, Vol. 41, No. 10, pp. 1738-1749.
17. Park S.-G., Choi Y., Oh Y.-J., Jeong K.-H. Terahertz photoconductive antenna with metal nanoislands, *Optics Express*, 2012, Vol. 20, No. 23, pp. 25530-25535.
18. Tanoto H., Teng J., Wu Q., et al. Nano-antenna in a photoconductive photomixer for highly efficient continuous-wave terahertz emission, *Scientific Reports*, 2013, Vol. 3, pp. 2824.
19. Polat E.O., Mercier G., Nikitskiy I., et al. Flexible graphene photodetectors for wearable fitness monitoring, *Science Advances*, 2019, Vol. 5, No. 9.
20. Balanis C.A. Antenna Theory: Analysis and Design. 4th ed. Hoboken: John Wiley & Sons, 2016, 1112 p.
21. Antipov S.A., Ishchenko E.A., Kostrova V.N. i dr. Issledovanie vliyaniya rasstoyaniya mezhdu antennami v MIMO-antennoy reshetke dlya setey pyatogo pokoleniya [Study of the influence of antenna spacing in a MIMO antenna array for fifth-generation networks], *Vestnik Voronezhskogo gosudarstvennogo tekhnicheskogo universiteta* [Bulletin of Voronezh State Technical University], 2020, Vol. 16, No. 4, pp. 91-96.
22. Sazonov D.M. Antenny i ustroystva SVCh [Microwave Antennas and Devices]. Moscow: Vysshaya shkola, 1988, 432 p.
23. Tian R., Lau B.K., Ying Z. Multiplexing efficiency of MIMO antennas, *IEEE Antennas and Wireless Propagation Letters*, 2011, Vol. 10, pp. 183-186.
24. Lamri I.E., Mansoul A., Nakmouche M.F., Belattar M. Design of Novel UWB 4-element MIMO Microstrip Patch Antenna for Sub-6 GHz 5G Applications, *International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)*, 2021, pp. 7-12.

Игнатович Владислав Олегович – Южный федеральный университет; e-mail: vignatovich@sfedu.ru; г. Таганрог, Россия; тел.: +79185530267; преподаватель Научно-образовательного центра «Перспективные решения в образовании».

Кисель Наталья Николаевна – Южный федеральный университет; e-mail: nnkisel@sfedu.ru; г. Таганрог, Россия; тел.: +78632184000, доб. 30020; к.т.н.; доцент кафедры антенн и радиопередающих устройств.

Ignatovich Vladislav Olegovich – Southern Federal University; e-mail: vignatovich@sfedu.ru; Taganrog, Russia; phone: +7918553-0267; teacher, Scientific and Educational Center "Perspective Decisions in Education".

Kisel Natalia Nikolaevna – Southern Federal University; e-mail: nnkisel@sfedu.ru; Taganrog, Russia; phone: +78632184000, add. 30020; cand. of eng. sc.; assistant professor, Department of Antennas and Radio Transmitting Devices.

А.В. Саенко, К.А. Козюменко, И.А. Шиховцов, Р.В. Томинов, В.А. Смирнов**ИССЛЕДОВАНИЕ РЕЗИСТИВНОГО ПЕРЕКЛЮЧЕНИЯ ПРОЗРАЧНЫХ
МЕМРИСТИВНЫХ СТРУКТУР НА ОСНОВЕ ОКСИДА ЦИНКА
ДЛЯ МАШИННОГО ЗРЕНИЯ РОБОТОТЕХНИЧЕСКИХ СИСТЕМ**

Развитие нейроморфных систем машинного зрения для робототехнических комплексов требует создания прозрачных мемристивных структур, сочетающих оптическую прозрачность, стабильное биполярное резистивное переключение и совместимость с технологией кроссбар-массивов. Ключевой проблемой является установление закономерностей влияния режимов осаждения тонких пленок ZnO на их структурные и электрофизические свойства, которые определяют характеристики мемристивных структур. Целью работы являлось определение оптимальной мощности ВЧ магнетронного распыления керамической мишени ZnO, обеспечивающей формирование прозрачных мемристивных структур ITO/ZnO/ITO со стабильным резистивным переключением, и создание на их основе кроссбар-массива. Тонкие пленки ZnO осаждались методом ВЧ магнетронного распыления при мощности от 25 до 100 Вт. Проведены структурные (РЭМ, АСМ) и электрофизические (эффект Холла) исследования полученных пленок ZnO. Изготовлены методами магнетронного распыления и литографии прозрачные мемристивные структуры ITO/ZnO/ITO и кроссбар-массив из 16 структур с размером ячеек 2000×2000 нм на стеклянных подложках, а также измерены их вольтамперные характеристики. Показано, что возрастание мощности магнетронного распыления от 25 до 100 Вт приводит к увеличению размера зерен от 12,8 до 35,7 нм и шероховатости поверхности пленок ZnO от 2,8 до 11,4 нм. Получено, что при мощности распыления 75 Вт концентрация носителей заряда в пленках ZnO имеет максимальное значение $2,7 \cdot 10^{15} \text{ см}^{-3}$, что необходимо для стабильного резистивного переключения структуры. Показано, что полученные мемристивные структуры ITO/ZnO/ITO проявляют стабильное биполярное переключение в течение 1000 циклов между состояниями $\text{HRS} = 537,4 \pm 26,7 \text{ Ом}$ и $\text{LRS} = 291,4 \pm 38,5 \text{ Ом}$ (отношение $\text{HRS/LRS} \sim 1,8$). Изготовленный прозрачный кроссбар-массив показал стабильное резистивное переключение в течение 20000 циклов ($\text{LRS} = 13,8 \pm 1,4 \text{ кОм}$, $\text{HRS} = 34,8 \pm 2,6 \text{ кОм}$, отношение $\text{HRS/LRS} \sim 2,5$). Полученные результаты могут быть использованы при разработке технологических процессов изготовления прозрачных мемристивных кроссбаров для нейроморфных структур машинного зрения робототехнических систем.

Наноматериалы; тонкие пленки ZnO; ВЧ магнетронное распыление; структурные свойства; нейроморфные системы; машинное зрение; прозрачный мемристоп; резистивное переключение.

A.V. Saenko, K.A. Kozyumenko, I.A. Shikhovtsov, R.V. Tominov, V.A. Smirnov**STUDY OF RESISTIVE SWITCHING OF TRANSPARENT ZINC OXIDE
MEMRISTIVE STRUCTURES FOR MACHINE VISION OF ROBOTIC SYSTEMS**

The development of neuromorphic machine vision systems for robotic systems requires the creation of transparent memristive structures that combine optical transparency, stable bipolar resistive switching, and compatibility with crossbar array technology. A key challenge is to establish patterns in the influence of ZnO thin film deposition modes on their structural and electrical properties, which determine the characteristics of memristive structures. The aim of this study was to determine the optimal RF magnetron sputtering power for a ZnO ceramic target, ensuring the formation of transparent ITO/ZnO/ITO memristive structures with stable resistive switching, and to create a crossbar array based on these structures. ZnO thin films were deposited using RF magnetron sputtering at powers ranging from 25 to 100 W. Structural (SEM, AFM) and electrical (Hall effect) studies of the resulting ZnO films were conducted. Transparent ITO/ZnO/ITO memristive structures and a crossbar array of 16 structures with a cell size of $2000 \times 2000 \text{ nm}$ were fabricated on glass substrates using magnetron sputtering and lithography, and their current-voltage characteristics were measured. Increasing the magnetron sputtering power from 25 to 100 W resulted in an increase in the grain size from 12,8 to 35,7 nm and in the surface roughness of the ZnO films from 2,8 to 11,4 nm. At a sputtering power of 75 W, the charge carrier concentration in the ZnO films reached a maximum value of $2.7 \times 10^{15} \text{ cm}^{-3}$, which is necessary for stable resistive switching of the structure. The obtained ITO/ZnO/ITO memristive structures were shown to exhibit stable bipolar switching for 1000 cycles between the states $\text{HRS} = 537,4 \pm 26,7 \text{ Ohm}$ and $\text{LRS} = 291,4 \pm 38,5 \text{ Ohm}$ (HRS/LRS ratio $\sim 1,8$). The fabricated transparent crossbar array showed stable resistive switching for 20000 cycles

($LRS = 13,8 \pm 1,4 \text{ k}\Omega$, $HRS = 34,8 \pm 2,6 \text{ k}\Omega$, HRS/LRS ratio $\sim 2,5$). The obtained results can be used in the development of technological processes for the fabrication of transparent memristor crossbars for neuromorphic structures of machine vision in robotic systems.

Nanomaterials; ZnO thin films; RF magnetron sputtering; structural properties; neuromorphic systems; machine vision; transparent memristor; resistive switching.

Введение. Мемристор является перспективным устройством энергонезависимой резистивной памяти благодаря таким достоинствам, как высокое быстродействие, длительное хранение данных, низкое энергопотребление и многоуровневое поведение [1–3]. Мемристор способен изменять свое сопротивление в зависимости от приложенного к нему напряжения и тем самым имитировать роль синапса в нервной системе, что открывает широкие возможности для разработки нейроморфных электронных устройств, таких как универсальные запоминающие устройства (ReRAM), вычислительные системы (нейрокомпьютеры), а также биосенсоры и нейроинтерфейсы для робототехники [4–6]. Развитие прозрачных мемристоров связано с их применением в оптоэлектронных нейроморфных устройствах, в частности системах искусственного машинного зрения, где наиболее важна оптическая прозрачность мемристора и его синаптическое поведение, поскольку мемристор напрямую реагирует на оптическое воздействие, а также обладает памятью и может быть использован в системах обработкой сенсорных данных и визуальной информации в реальном времени [7, 8]. Прозрачные мемристоры могут быть также интегрированы в устройства прозрачной электроники, такие как тонкопленочные транзисторы и диоды. Кроме того, мемристоры с оптической бистабильностью могут найти широкое применение в создании функциональных элементов для оптоэлектроники (высокоскоростные оптические переключатели и модуляторы света) и дисплеев интегрированных, например, в окна и лобовые стекла автомобилей [9–11]. Для создания эффективных прозрачных мемристоров необходимы оптически прозрачные и электрически переключаемые оксидные материалы, а также прозрачные верхний и нижний электроды, такие как ИТО [3, 12].

В последнее время перспективным мемристивным наноматериалом выступает оксид цинка (ZnO), который является не токсичным, обладает подходящей шириной запрещенной зоны ($\sim 3,4 \text{ эВ}$) и уникальной комбинацией оптических и электрофизических свойств, которые во многом определяются концентрацией свободных носителей заряда, контролируемых кислородными вакансиями, а также проявляет синаптическое поведение, необходимое для создания прозрачных мемристивных структур [13–15].

При осаждении тонких пленок ZnO широко используются такие методы, как магнетронное распыление на постоянном токе и в высокочастотном режиме (ВЧ), химическое осаждение из газовой фазы, импульсное лазерное осаждение и золь-гель метод (центрифугирование) [16–18]. Метод магнетронного распыления имеет такие преимущества, как возможность точного управления параметрами растущей пленки в процессе осаждения, высокая стабильность процесса во времени и отсутствие нагрева подложки до высоких температур [4]. Обычно выделяют два метода магнетронного распыления оксидов металлов, которые включают распыление металлической мишени в среде реактивного газа (кислород) или керамической мишени в атмосфере инертного газа (аргон). При этом распыление керамической мишени характеризуется высокой стабильностью процесса и высоким качеством получаемых пленок, в отличие от реактивного распыления (низкая стабильность процесса и сложность его контроля), что может быть перспективным для создания мемристивных структур [12, 19].

Целью данной работы является исследование влияния мощности ВЧ магнетронного распыления керамической мишени при комнатной температуре в атмосфере аргона на структурные и электрофизические свойства пленок ZnO, формирование и исследование резистивного переключения прозрачных мемристивных структур ИТО/ZnO/ИТО, а также кроссбар-массива на их основе.

Методика эксперимента. Прозрачные мемристивные структуры ITO/ZnO/ITO формировались в три этапа методом магнетронного распыления на установке VSE-PVD-DESK-PRO (ООО «АкадемВак»). На первом этапе пленки ITO толщиной 200 нм (нижний электрод) осаждались на стеклянные подложки методом импульсного магнетронного распыления на постоянном токе с частотой 100 кГц при комнатной температуре в атмосфере аргона, мощности распыления 200 Вт и рабочем давлении $2 \cdot 10^{-3}$ мбар [20–22]. На втором этапе тонкие пленки ZnO осаждались на стеклянные подложки с прозрачным проводящим слоем ITO методом ВЧ магнетронного распыления керамической мишени ZnO чистотой 99,99 % при комнатной температуре в атмосфере аргона [1, 15]. Мощность распыления изменялась от 25 до 100 Вт, а рабочее давление аргона устанавливалось на уровне $5 \cdot 10^{-3}$ мбар. Время осаждения изменялось от 4 до 15 мин., при котором толщина полученных пленок ZnO составляла от 56 до 69 нм. На третьем этапе верхний ITO электрод толщиной 150 нм и диаметром 550 мкм осаждался на поверхность пленки ZnO также методом магнетронного распыления на постоянном токе с частотой 100 кГц при комнатной температуре в атмосфере аргона.

Изготовление прозрачного кроссбара 4×4 мемристивных структур проводилось также в три этапа. На первом этапе на стеклянные подложки наносилась прозрачная пленка ITO методом магнетронного распыления в импульсном режиме, на поверхности которой затем проводилась фотолитография и формировались нижние электроды. На втором этапе методом центрифугирования наносилась пленка фоторезиста ФП-383, которая была экспонирована через фотошаблон с массивом ячеек мемристивных структур и проявлена в 5% растворе КОН. Далее на полученную структуру наносилась пленка ZnO методом ВЧ магнетронного распыления при мощности 75 Вт. Затем после взрывной литографии пленки ZnO были получены ячейки мемристивных структур. На третьем этапе формировались верхние ITO контакты методами магнетронного распыления на постоянном токе с частотой 100 кГц и взрывной литографии. Для обеспечения возможности проведения зондовых электрических измерений каждый электрод был изготовлен с контактной площадкой.

Морфология поверхности пленок ZnO исследовалась с помощью растрового электронного микроскопа Nova Nanolab 600 (FEI Company) и атомно-силового микроскопа (АСМ) в полуконтактном режиме на сканирующем зондовом микроскопе NTEGRA (НТ-МДТ). Электрофизические параметры пленок ZnO измерялись на установке Есоріа HMS-3000 (Есоріа Со.). Коэффициент пропускания мемристивных структур в ультрафиолетовом и видимом спектральном диапазоне (350–900 нм) определялся с помощью спектрофотометра UV-VIS Evolution-300 (Thermo). Измерения вольтамперных характеристик мемристивных структур проводилось при комнатной температуре на воздухе с использованием системы измерений параметров полупроводников Keithley 4200-SCS и установки субмикронного зондирования ЭМ-6070А с вольфрамовыми зондами. Напряжение подавалось на верхний электрод мемристивной структуры, а нижний электрод был заземлен [10, 11]. Для исследования электрических параметров изготовленного кроссбара, для каждой из 16 структур были измерены 1000 вольтамперных характеристик в режиме биполярной развертки от -3 В до 3 В. Также было проведено исследование зависимости сопротивлений HRS и LRS от количества цикла переключения. Для этого на каждую структуру подавалось 20000 циклов переключения при напряжении чтения 0,4 В. Существует несколько способов минимизировать влияние токов утечки в пассивных кроссбарах [23, 24]: путем подачи напряжений равных $U/2$ В и $U/3$ В на неиспользуемые электроды и путем заземления неиспользуемых электродов. В данном исследовании для минимизации влияния токов утечки все неиспользуемые электроды были заземлены во время электрических измерений. Кроме того, следует отметить, что низкое отношение HRS/LRS приводит к уменьшению влияния токов утечки в пассивных кроссбарах.

Результаты и обсуждение. На рис. 1 представлены РЭМ-изображения поверхности пленок ZnO, полученных при мощности ВЧ магнетронного распыления от 25 до 100 Вт, и их поперечный скол толщиной 60 нм. Анализ РЭМ-изображений показал, что пленки

ZnO, полученные при мощности ВЧ магнетронного распыления от 25 до 75 Вт, имеют однородную зернистую морфологию поверхности без существенных дефектов со средним размером зерен от 12,8 до 28,6 нм (рис. 1,а-в). При этом пленки ZnO, полученные при мощности распыления 100 Вт, имеют существенные неоднородности, связанные с изменяющимся размером зерен от 14,3 до 53,5 нм и средний размер зерен 35,7 нм (рис. 1,г). На поперечном срезе всех пленок ZnO видна столбчатая структура с направлением перпендикулярным к плоскости подложки, которая часто наблюдается при низкотемпературном осаждении пленок оксидов и является следствием малой подвижности осаждаемых частиц на поверхности подложки [17–19].

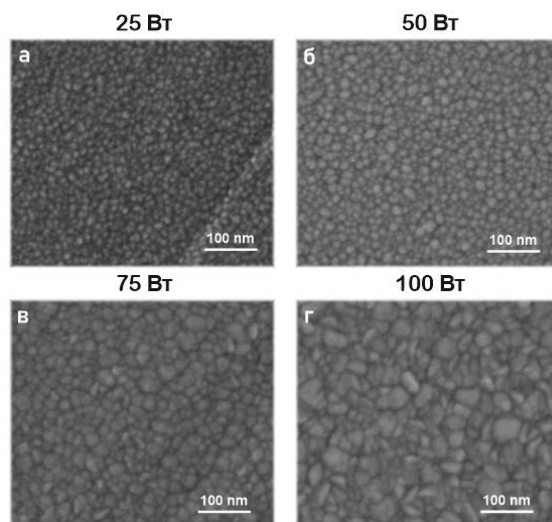


Рис. 1. РЭМ-изображения поверхности пленок ZnO, полученных при мощности ВЧ магнетронного распыления от 25 до 100 Вт

На рис. 2 представлены АСМ-изображения поверхности пленок ZnO, полученных при мощности ВЧ магнетронного распыления от 25 до 100 Вт. Анализ АСМ-изображений показал, что пленки ZnO имеют относительно гладкую поверхность при мощности магнетронного распыления от 25 до 75 Вт. Так при мощности распыления 25 Вт шероховатость поверхности пленки составляет 2,8 нм, при мощности распыления 50 Вт – 5,1 нм, а при мощности распыления 75 Вт – порядка 6,3 нм (рис. 2, а-в). При мощности распыления 100 Вт наблюдается увеличение шероховатости поверхности до 11,4 нм за счет присутствия в пленке зерен большего размера (40-60 нм) (рис. 2, г). Возрастание мощности ВЧ магнетронного распыления придает осаждаемым атомам необходимую кинетическую энергию для роста зерен большего размера, что приводит к возрастанию шероховатости поверхности и может приводить к снижению стабильности резистивного переключения [2, 4].

Кроме того, при низкой мощности (25 Вт) величина шероховатости определяется аморфноподобной структурой полученных образцов: образование более мелких полостей или каналов способствует формированию гладкой поверхности с меньшей шероховатостью. Увеличение мощности до 100 Вт приводит к изменению структуры пленок, образованию более глубоких каналов и увеличению шероховатости [23]. Таким образом, зависимость шероховатости от мощности распыления тесно связана с особенностями кристаллической структуры, размером зерна и распределением дефектов на границах зерен. Следует отметить, что температура подложки и поток газа также могут быть изменены для управления шероховатостью поверхности. Ожидается, что увеличение температуры подложки во время магнетронного распыления приведет к формированию более мелкозернистых пленок, однако этот подход может быть использован в ограниченной степени в рамках данной работы, поскольку нагрев подложки во время осаждения пленки ZnO может ухудшить параметры прозрачного проводящего слоя ИТО.

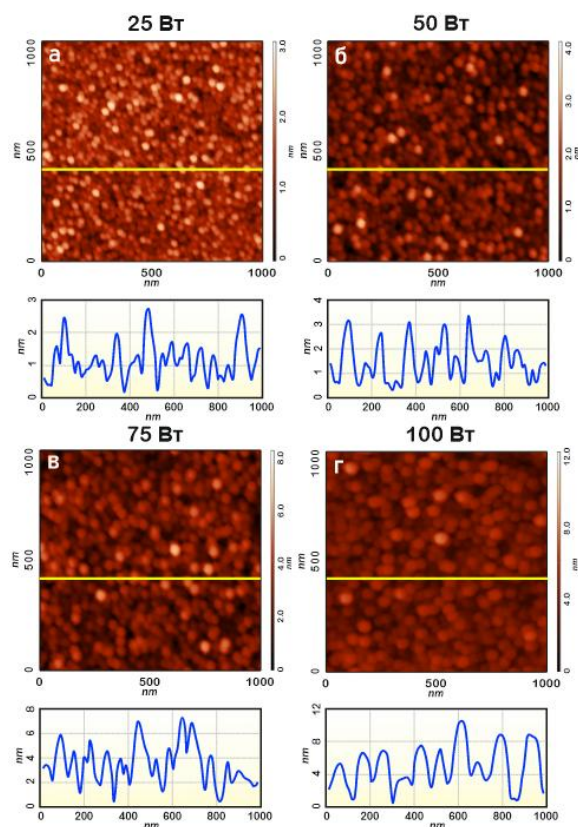


Рис. 2. АСМ-изображения поверхности пленок ZnO, полученных при мощности ВЧ магнетронного распыления от 25 до 100 Вт

На рис. 3 представлены зависимости концентрации и подвижности носителей заряда в пленках ZnO от мощности ВЧ магнетронного распыления.

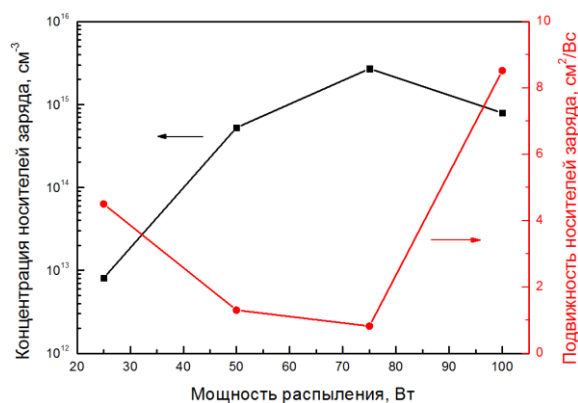


Рис. 3. Электрофизические свойства пленок ZnO, полученных при мощности ВЧ магнетронного распыления от 25 до 100 Вт

Полученные пленки ZnO имеют n-тип проводимости и относительно высокое удельное сопротивление, которое составляло порядка 10^3 - 10^4 Ом·см. Концентрация носителей заряда в пленках ZnO существенно изменяется от $8,1 \cdot 10^{12} \text{ см}^{-3}$ до $2,7 \cdot 10^{15} \text{ см}^{-3}$ при увеличении мощности ВЧ магнетронного распыления от 25 до 75 Вт. Данные изменения концентрации носителей заряда можно связать с кристалличностью пленок ZnO, в

частности размером зерен, а также рассеянием на границах зерен и дефектах [6, 17]. Подвижность носителей заряда изменяется не так значительно и имеет противоположный характер. Таким образом, при мощности распыления 75 Вт в пленке ZnO наблюдается наибольшая концентрация носителей заряда, которая определяется низкими энергиями образования внутренних дефектов нестехиометрии – анионными (кислородными) вакансиями, что необходимо для стабильного резистивного переключения мемристивных структур на основе филаментарного механизма.

Средний коэффициент пропускания в видимом спектре (длина волны 400-800 нм) составил около 86 %, что соответствует высокой оптической прозрачности в сочетании с перспективной резистивной памятью для нейроморфных структур машинного зрения, робототехники и систем искусственного интеллекта.

Для исследования резистивного переключения мемристивных структур ITO/ZnO/ITO проведено измерение их электрических характеристик. На рис. 4 представлены экспериментальная вольтамперная характеристика, показывающая поведение биполярного резистивного переключения, зависимость сопротивления от количества циклов переключения и кумулятивная вероятность для мемристивной структуры, в которой пленка ZnO осаждалась при мощности распыления 75 Вт.

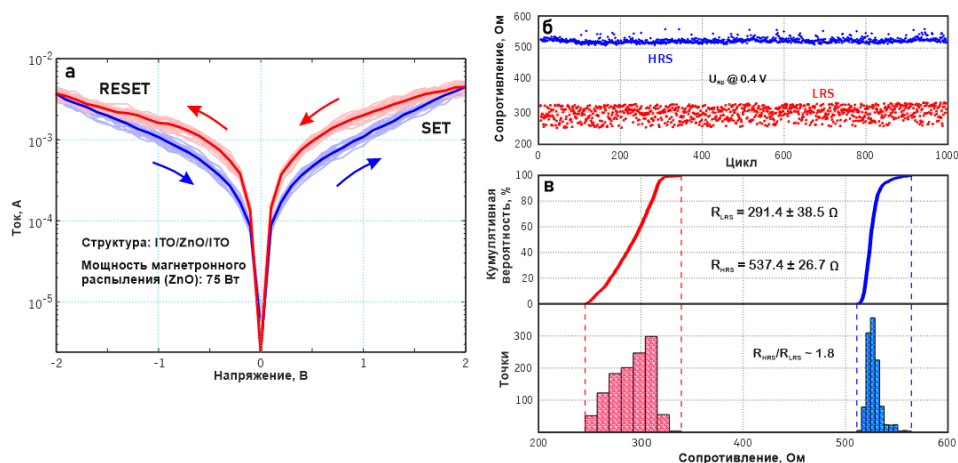


Рис. 4. Экспериментальные исследования резистивного переключения прозрачной мемристивной структуры ITO/ZnO/ITO

♦ Измерение вольтамперной характеристики мемристивной структуры осуществлялось при изменении приложенного к верхнему электроду напряжения в последовательности $0 \text{ В} \rightarrow 2 \text{ В} \rightarrow 0 \text{ В} \rightarrow -2 \text{ В} \rightarrow 0 \text{ В}$, что показано в виде стрелок на рис. 4,а [13, 23]. На рис. 4,б представлена зависимость высокого (HRS) и низкого (LRS) сопротивлений пленки ZnO от количества циклов переключения. При увеличении приложенного напряжения от 0 В до 2 В мемристивная структура ITO/ZnO/ITO постепенно переходит из состояния HRS в LRS и процесс установки завершается при 2 В (VSET). В дальнейшем мемристивная структура будет сохранять состояние LRS до тех пор, пока приложенное напряжение не снизится до -2 В (VRESET). Следовательно, при VRESET происходит процесс сброса и мемристор переходит из состояния LRS в HRS, реализуя биполярное резистивное переключение. Таким образом, показано изменение сопротивления из состояния HRS в LRS, которое происходит при $2 \pm 0,1 \text{ В}$, а из LRS в HRS при $-2 \pm 0,1 \text{ В}$, при этом ток переключения составляет порядка 5 мА. Получено, что за 1000 циклов переключения HRS и LRS составляют $537,4 \pm 26,7 \text{ Ом}$ и $291,4 \pm 38,5 \text{ Ом}$, соответственно. При этом отношение сопротивлений в состояниях HRS/LRS составляет $\sim 1,8$ при напряжении чтения 0,4 В (рис. 4,в). В целом пленка ZnO является диэлектриком, и для того, чтобы она начала проявлять эффект резистивного переключения, необходимо создать в ее объеме определенное количество кислородных вакансий. Этого можно до-

биться несколькими способами: электроформовкой, подбором режимов распыления для создания пленок с определенной стехиометрией, а также применением отжига после напыления [17, 21]. В данной работе с помощью магнетронного распыления найдены значения давления в камере и параметров мощности распыления, при которых пленки ZnO начинают проявлять резистивное переключение. Таким образом, наличие резистивного переключения в полученных пленках ZnO связано с условиями распыления.

Для практического применения исследуемых в данной работе прозрачных мемристоров, в частности для эмуляции на их основе биологических синапсов в нейроморфных системах машинного зрения, необходимо разработать и изготовить массивы элементов, которые будут электрически соединены между собой. Считается, что для физической реализации биологических синапсов наиболее подходящими являются мемристивные структуры, изготовленные по архитектуре кроссбара, так как они способны эмулировать вычислительные примитивы клеток нейронов [23]. Исходя из вышесказанного, был разработан и изготовлен прозрачный кроссбар из 16 мемристивных структур ITO/ZnO/ITO с размером 2000×2000 нм каждая (рис. 5).

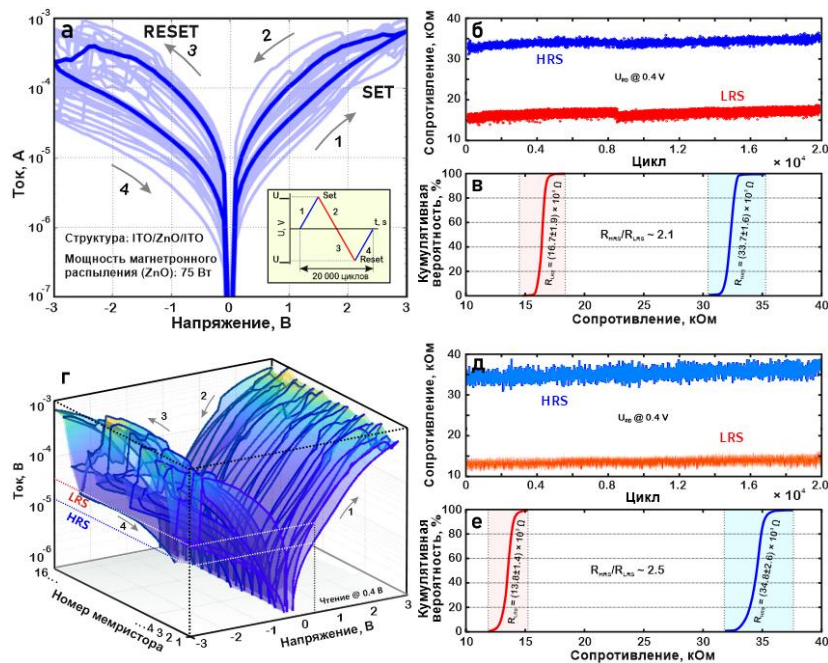


Рис. 5. Исследование резистивного переключения кроссбар массива из 16 прозрачных мемристивных структур

Анализ полученных вольтамперных характеристик показал, что изготовленные мемристивные структуры обладают стабильным резистивным переключением (рис. 5,а). Ток в состоянии HRS составляет около 15 мкА при напряжении 0,4 В, а в состоянии низкого сопротивления LRS 32 мкА. Аналогичные результаты были получены для остальных мемристивных структур кроссбара (рис. 5,г). Испытание на выносливость показало, что за 20000 циклов переключения мемристивные структуры переключаются между $LRS = 16,7 \pm 1,9$ кОм и $HRS = 33,7 \pm 1,6$ кОм для одного устройства (рис. 5,б), а также между $LRS = 13,8 \pm 1,4$ кОм и $HRS = 34,8 \pm 2,6$ кОм для всех устройств (рис. 5,д), в то время как отношение сопротивлений HRS/LRS составляет $\sim 2,1$ и $2,5$ соответственно (рис. 5,в,е). Данное различие может быть объяснено влиянием поверхностного рельефа при одной и той же шероховатости и толщине пленки ZnO на эффект резистивного переключения, что приводит к флуктуациям концентрации и профиля распределения дефектов в объеме пленки ZnO для каждой мемристивной структуры ITO/ZnO/ITO. Изготов-

ленный кроссбар мемристивных структур на основе ИТО/ZnO/ИТО может быть использован в перспективных приложениях для робототехники и систем искусственного интеллекта, в частности в зрительных системах, а также в системах для распознавания образов.

Заключение. В результате проведенных исследований установлены закономерности влияния мощности ВЧ магнетронного распыления керамической мишени ZnO при комнатной температуре в атмосфере аргона на структурные и электрофизические свойства тонких пленок ZnO. Показано, что увеличение мощности распыления от 25 до 100 Вт приводит к росту размера зерен от 12,8 до 35,7 нм и повышению шероховатости поверхности от 2,8 до 11,4 нм. Определена оптимальная мощность магнетронного распыления (75 Вт), при которой пленки ZnO имеет однородную нанокристаллическую структуру со средним размером зерен 28,6 нм, шероховатостью поверхности 6,3 нм и наибольшей концентрацией носителей заряда $2,7 \cdot 10^{15} \text{ см}^{-3}$, что необходимо для стабильного резистивного переключения мемристивной структуры на основе филаментарного механизма. Показана возможность создания прозрачных мемристивных структур ИТО/ZnO/ИТО и кроссбар-массив из 16 структур с размером ячеек $2000 \times 2000 \text{ нм}$ на стеклянных подложках. Получено, что при мощности осаждения пленок ZnO 75 Вт мемристивные структуры показывают стабильное биполярное резистивное переключение в течение 1000 циклов между сопротивлениями $\text{HRS} = 537,4 \pm 26,7 \text{ Ом}$ и $\text{LRS} = 291,4 \pm 38,5 \text{ Ом}$ при отношении $\text{HRS/LRS} \sim 1,8$. Установлено, что все структуры кроссбар-массива также показывают стабильное резистивное переключение в течение 20000 циклов между $\text{LRS} = 13,8 \pm 1,4 \text{ кОм}$ и $\text{HRS} = 34,8 \pm 2,6 \text{ кОм}$ при отношении $\text{HRS/LRS} \sim 2,5$. Получено, что коэффициент оптического пропускания мемристивных структур в видимом спектре составляет $\sim 86 \%$, что подтверждает их высокую прозрачность. Представленные результаты могут быть использованы при разработке технологических процессов создания прозрачных мемристивных кроссбаров для нейроморфных систем машинного зрения, энергонезависимой резистивной памяти и оптоэлектронных структур робототехнических комплексов.

Исследование выполнено за счёт гранта Российского научного фонда № 25-19-00809, <https://rscf.ru/project/25-19-00809/> в Южном федеральном университете.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Jing Yang, Lingxiang Hu, Liufeng Shen, Jingrui Wang, Peihong Cheng, Huanming Lu, Fei Zhuge, Zhizhen Ye. Optically driven intelligent computing with ZnO memristor, *Fundamental Research*, 2024, Vol. 4, pp. 158-166.
2. Usman Bature Isyaku, Mohd Haris Bin Md Khir, I. Md Nawi, M.A. Zakariya, Furqan Zahoor. ZnO Based Resistive Random Access Memory Device: A Prospective Multifunctional Next-Generation Memory, *IEEE Access*, 2021, Vol. 9, pp. 105012.
3. P. Praveen, T. Priya Rose, K.J. Saji. Top electrode dependent resistive switching in M/ZnO/ITO memristors, M = Al, ITO, Cu, and Au, *Microelectronics Journal*, 2022, Vol. 121, pp. 105388.
4. Saenko A.V., Tominov R.V., Jityaev I.L., Vakulov Z.E., Avilov V.I., Polupanov N.V., Smirnov V.A. Transparent Zinc Oxide Memristor Structures: Magnetron Sputtering of Thin Films, Resistive Switching Investigation, and Crossbar Array Fabrication, *Nanomaterials*, 2024, Vol. 14, pp. 1901.
5. Cristian L. Teran, Jorge A. Calderon, Heiddy P. Quiroz, A. Dussan. Optical properties and bipolar resistive switching of ZnO thin films deposited via DC magnetron sputtering, *Chinese Journal of Physics*, 2021, Vol. 74, pp. 1-8.
6. Chander Prakash, Lovi Raj Gupta, Amrinder Mehta, Hitesh Vasudev, Roman Tominov, Ekaterina Korman, Alexander Fedotov, Vladimir Smirnov, Kavindra Kumar Kesari. Computing of neuromorphic materials: an emerging approach for bioengineering solutions, *Materials Advances*, 2023, Vol. 4, pp. 5882-5919.
7. Yuxin Shi, Yanna Zhang, Guoqiang Li. Recent progress in transparent memristors, *Journal of Physics D: Applied Physics*, 2023, Vol. 56, pp. 313001.
8. Huiling Zhang, Ruping Liu, Huiqing Zhao, Zhicheng Sun, Zilong Liu, Liang He, Ye Li. Research Progress of Biomimetic Memristor Flexible Synapse, *Coatings*, 2022, Vol. 12, pp. 21.
9. Ping-Xing Chen, Debashis Panda, Tseung-Yuen Tseng. All oxide based flexible multi-folded invisible synapse as vision photoreceptor, *Scientific Reports*, 2023, Vol. 13, pp. 1454.

10. Raveendra Kiran M., Hidayath Ulla, Satyanarayan M.N., Umesh G. Effects of annealing temperature on the resistance switching behaviour of solution-processed ZnO thin films, *Superlattices and Microstructures*, 2020, Vol. 148, 106718.
11. Asutosh Patnaik, Srikant Kumar Mohanty, Narayan Sahoo, Debashis Panda. Effect of oxygen concentration in ZnO-based transparent flexible memristor synapse, *Journal of Materials Science: Materials in Electronics*, 2023, Vol. 34, pp. 1406.
12. Hongxia Li, Wei Dong, Xin Wu, Junhua Xi, Zhenguo Ji. Resistive switching characteristics of ZnO/a-TiO₂ bilayer film fabricated on PET/ITO transparent and flexible substrates, *Materials Research Bulletin*, 2016, Vol. 84, pp. 449-454.
13. Tominov R.V., Vakulov Z.E., Avilov V.I., Shikhovtsov I.A., Varganov V.I., Kazantsev V.B., Lovi Raj Gupta, Chander Prakash, Smirnov V.A. Approaches for Memristive Structures Using Scratching Probe Nanolithography: Towards Neuromorphic Applications, *Nanomaterials*, 2023, Vol. 13, pp. 1583.
14. Abduev A.Kh., Akhmedov A.K., Asvarov A.Sh., Muslimov A.E., Kanevsky V.M. ZnO-based transparent conductive layers obtained by the magnetron sputtering of a composite cermet ZnO:Ga–Zn target: part 2, *Journal of Surface Investigation. X-Ray, Synchrotron and Neutron Techniques*, 2021, Vol. 15, pp. 121-127.
15. Sobia Ali Khan, Geun Ho Lee, Chandreswar Mahata, Muhammad Ismail, Hyungjin Kim, Sungjun Kim. Bipolar and Complementary Resistive Switching Characteristics and Neuromorphic System Simulation in a Pt/ZnO/TiN Synaptic Device, *Nanomaterials*, 2021, Vol. 11, pp. 315.
16. Kavindra Kandpal, Jitendra Singh, Navneet Gupta, Chandra Shekhar. Effect of thickness on the properties of ZnO thin films prepared by reactive RF sputtering, *Journal of Materials Science: Materials in Electronics*, 2018, Vol. 29, pp. 14501-14507.
17. Zhiqiang Yu, Jinhao Jia, Xinru Qu, Qingcheng Wang, Wenbo Kang, Baosheng Liu, Qingquan Xiao, Tinghong Gao, Qian Xie. Tunable Resistive Switching Behaviors and Mechanism of the W/ZnO/ITO Memory Cell // *Molecules*. – 2023. – Vol. 28. – P. 5313.
18. Paulina Kaim, Krzysztof Lukaszewicz, Marek Szindler, Magdalena M. Szindler, Marcin Basiaga, Barbara Hajduk. The influence of magnetron sputtering process temperature on ZnO thin-film properties, *Coatings*, 2021, Vol. 11, pp. 1507.
19. Ziyu Lv, Yan Wang, Jingrui Chen, Junjie Wang, Ye Zhou, Su-Ting Han. Semiconductor Quantum Dots for Memories and Neuromorphic Computing Systems, *Chemical Reviews Journal*, 2020, Vol. 120, pp. 3941-4006.
20. Ziyu Lv, Shirui Zhu, Yan Wang, Yanyun Ren, Mingtao Luo, Hanning Wang, Guohua Zhang, Yongbiao Zhai, Shilong Zhao, Ye Zhou, Minghao Jiang, Yan-Bing Leng, Su-Ting Han. Development of Bio-Voltage Operated Humidity-Sensory Neurons Comprising Self-Assembled Peptide Memristors, *Advanced Materials*, 2024, Vol. 36, pp. 2405145.
21. Tominov R.V., Vakulov Z. E., Polupanov N.V., Saenko A.V., Avilov V.I., Ageev O.A., Smirnov V.A. Nanoscale-resistive switching in forming-free zinc oxide memristive structures, *Nanomaterials*, 2022, Vol. 12, pp. 455.
22. Saenko A.V., Vakulov Z.E., Klimin V.S., Bilyk G.E., Malyukov S.P. Effect of Magnetron Sputtering Power on ITO Film Deposition at Room Temperature, *Russian Microelectronics*, 2023, Vol. 23, pp. 297-302.
23. Serb A., Bill J., Khiat A., Berdan R., Legenstein R., Prodromakis T. Unsupervised learning in probabilistic neural networks with multi-state metal-oxide memristive synapses, *Nature communications*, 2016, Vol. 7, pp. 12611.

Саенко Александр Викторович – Южный федеральный университет; e-mail: avsaenko@sfnu.ru; г. Таганрог, Россия; тел. +78634371767; к.т.н.; доцент.

Козюменко Константин Алексеевич – Южный федеральный университет; e-mail: koziumenko@sfnu.ru; г. Таганрог, Россия; тел. +78634371767; лаборант-исследователь.

Шиховцов Иван Андреевич – Южный федеральный университет; e-mail: shihovcov@sfnu.ru; г. Таганрог, Россия; тел.: +78634371767; аспирант.

Томинов Роман Викторович – Южный федеральный университет; e-mail: tominov@sfnu.ru; г. Таганрог, Россия; тел.: +78634371767; к.т.н.; доцент.

Смирнов Владимир Александрович – Южный федеральный университет; e-mail: vasmimov@sfnu.ru; г. Таганрог, Россия; тел. +78634371767; к.т.н.; зав. кафедрой.

Saenko Aleksandr Viktorovich – Southern Federal University; e-mail: avsaenko@sfedu.ru; Taganrog, Russia; phone: +78634371767; cand. of eng. sc., associate professor.

Kozyumenko Konstantin Alekseevich – Southern Federal University; e-mail: koziumenko@sfedu.ru; Taganrog, Russia; phone: +78634371767; research lab assistant.

Shikhovtsov Ivan Andreevich – Southern Federal University; e-mail: shihovcov@sfedu.ru; Taganrog, Russia; phone: +78634371767; postgraduate student.

Tominov Roman Viktorovich – Southern Federal University; e-mail: tominov@sfedu.ru; Taganrog, Russia; phone: +78634371767; cand. of eng. sc., associate professor.

Smirnov Vladimir Aleksandrovich – Southern Federal University; e-mail: vasmirnov@sfedu.ru; Taganrog, Russia; phone: +78634371767; cand. of eng. sc.; head of department.

Раздел IV. Машинное обучение и нейронные сети

УДК 004.032.26

DOI 10.18522/2311-3103-2026-1-235-249

Б.В. Костров, С.И. Бабаев, А.И. Ефимов, В.Ю. Тарасова

РАЗРАБОТКА ГИБРИДНОЙ АРХИТЕКТУРЫ СВЁРТОЧНОЙ НЕЙРОННОЙ СЕТИ С ИСПОЛЬЗОВАНИЕМ СПЕКТРАЛЬНЫХ ПРЕОБРАЗОВАНИЙ

Предлагается архитектура гибридной свёрточной нейронной сети, сочетающей спектральные и пространственные слои, а также новый метод подвыборки (WalsPooling). Разработанная система применяется для оценки географической близости пары изображений на основе их визуального сходства. Пара изображений получена с разных сенсоров и определению их визуального сходства осложняется условиями съёмки с разным масштабом, углом наклона сенсора. На основе фрагмента съёмки на малой высоте выполняется поиск в базе данных изображений подстилающей поверхности. Поиск осуществляется в окрестности заданного маршрута на основе вектора признаков изображений, который формируется на последнем слое свёрточной нейронной сети. Система использует сиамскую архитектуру, так как на вход необходимо подавать пару изображений. Актуальность задачи обусловлена необходимостью обеспечения навигации БПЛА в условиях отсутствия или ненадёжности GPS-сигнала. Также рассматривается подход к формированию набора данных и его предварительная обработка. Поиск в базе данных выполняется в окрестности маршрута, что сокращает вычислительные затраты. Эксперименты включают анализ применимости предложенных слоёв (WalsPooling) и сравнение с традиционными методами подвыборки. Также в работе представлен метод линейной аппроксимации с обучаемыми параметрами для уменьшения размерности свёрточного слоя. Основное преимущество подхода – устойчивость к изменениям масштаба и угла съёмки за счёт комбинации спектральных и пространственных признаков. Результаты демонстрируют применимость метода для навигации БПЛА в условиях потери или ненадёжности GPS-сигнала. В результате эксперимента было получено, что использование восстановленных после спектрального преобразования изображений позволяет достичь наилучших результатов сходимости нейронной сети и среднеквадратической ошибки. Разработанная архитектура демонстрирует устойчивость к геометрическим и яркостным искажениям, а ее метрики качества (Precision = 0.728, Recall = 0.800, F1 = 0.872) подтверждает эффективность подхода для задач визуальной локализации на основе изображений из базы данных подстилающей поверхности.

Гибридная СНС; преобразование Уолша; визуальная локализация; сиамские сети; визуальный поиск; WalsPooling.

B.V. Kostrov, S.I. Babaev, A.I. Efimov, V.Yu. Tarasova

CONVOLUTIONAL NEURAL NETWORK HYBRID ARCHITECTURE DEVELOPMENT USING SPECTRAL TRANSFORMATIONS

The hybrid convolutional neural network architecture with combining spectral and spatial layers, as well as new methods of subsampling (WalsPooling) and convolution (ConvWals) are proposed. The developed system is used to geographical proximity assess of images pair based on their visual similarity. A pair of different sensors obtained images visual similarity determination is complicated by different scales and sensor tilt angles shooting conditions. Based on the low-altitude image fragment, a search in the database of underlying surface images is performed. The search is performed in the surrounding area of a given route based on the vector of image features, which is formed on the last layer of the convolutional neural network. The system uses the Siamese architecture, since a pair of images must be submitted to the input. The relevance of this problem stems from the need to ensure UAV navigation in the absence or unreliability of a GPS signal. The approach to data set formation and its preprocessing is also considered. The database search is

performed in the surrounding area of the route, which reduces computational costs. The experiments include an analysis of the applicability of the proposed layers (WalsPooling, ConvWals) and a comparison with traditional pooling and convolution methods. The paper also presents a linear approximation method with trainable parameters for reducing the dimensionality of the convolutional layer. The main advantage of the approach is its resistance to changes in the scale and angle of shooting due to a combination of spectral and spatial features. The results demonstrate the applicability of the method for UAV navigation in conditions of loss of GPS signal is lost or unreliable. The experiment demonstrated that using images reconstructed after spectral transformation yields the best neural network convergence and mean square error. The developed architecture demonstrates robustness to geometric and brightness distortions, and its quality metrics (Precision = 0.728, Recall = 0.800, F1 = 0.872) confirm the effectiveness of the approach for visual localization tasks based on images from a surface database.

Hybrid SNS; Walsh transformation; visual localization; Siamese networks; visual search; WalsPooling.

Введение. Данная работа посвящена разработке архитектуры гибридной свёрточной нейронной сети (СНС) с новыми типами слоёв и методами обработки изображений. В работе предлагается новая структура гибридной СНС (сочетание спектральных и стандартных слоёв). Стандартные слои СНС применяются для выделения локальных признаков, спектральные – для глобальных. Предложен новый метод подвыборки (пулинга) на основе линейной аппроксимации с обучаемым параметром (модель сама выбирает к какой точке аппроксимировать), который в сочетании с WalsPooling улучшает сохранение пространственно-спектральных признаков.

В данной работе описывается способ взаимодействия компонентов (сиамская сеть, среднеквадратическое отклонение), а также предварительная обработка данных, которая состоит в применении спектрального преобразования Уолша, удалении низкочастотной составляющей. Предлагается использовать сиамскую архитектуру, так как необходимо только по визуальному сходству пары изображений оценить степень их географической близости. Степень близости оценивается как среднеквадратическое отклонение векторов признаков.

Данная архитектура применяется в задаче локализации БПЛА в условиях отсутствия GPS-сигнала на основе визуального сходства пар аэрофотоснимков. На вход нейронной сети подаются два изображения: текущий снимок с БПЛА (получен в реальном времени во время полёта) и спутниковый снимок из базы данных подстилающей поверхности (геоприязанный). Известны маршрут полёта (стартовая и конечная точки) и GPS-координаты в начальный момент, но в процессе полёта сигнал GPS недоступен или ненадёжен. Цель работы оценить степень географической близости двух снимков на основе их визуального сходства, чтобы определить текущее положение БПЛА и при необходимости скорректировать траекторию. В данной работе рассматривается только задача сравнения изображений (оценка визуального сходства), а не полная навигационная система.

Обзор существующих решений. В работах [1–4] предлагается заменить классические методы подвыборки (операции уменьшения размерности) для сохранения большего объёма информации. В работе [1] применяется интерполяционный пулинг, учитывающий веса ближайших пикселей, что близко к предложенной линейной аппроксимации, однако кроме учета веса ближайших пикселей, также рассматривается обучаемый параметр для смещений (помимо весов еще выбирается к какому пикселю смещаться при аппроксимации). Предложенная линейная интерполяция с обучаемым параметром, так же как и в работах [1, 3, 4] направлена на минимизацию потерь важных признаков при уменьшении размерности. В работе [2] акцент сделан на ускорение вычислений. В работах [3, 4] используется вейвлет-пулинг, который также сохраняет структурную информацию за счёт спектрального разложения. Ключевым отличием предложенного WalsPooling, является то, что он применяется совместно с линейной аппроксимацией, что уникально и позволяет добиться сохранения пространственно-спектральных признаков. Вейвлет-пулинг [3, 4] использует другое спектральное представление, и в отличие от спектрального преобразования Уолша, он требует больше ресурсов и не адаптирован для сравнения изображений. Работы [1–4] тестируются на задачах классификации, с использованием набора данных

ImageNet [5], а не парном сравнении как в данном исследовании. Линейная аппроксимация имеет сходство с интерполяционным пулингом [1], но в данной работе смещение предлагается сделать обучаемым и для каждой соседней точки задать обучаемые веса (взвешенное суммирование). Предложенный метод имеет высокую устойчивость к искажениям, в отличие от рассмотренных работ.

В работах [6–8] применяются вейвлет-преобразования для уменьшения потери информации в СНС. В [6] вейвлеты встроены в нейронную сеть U-Net для задач восстановления изображений. В [7] вейвлеты используются для предобработки и ускорения СНС. В работе [8] используется вейвлет-преобразование Хаара для шумоподавления и выделения признаков. В работе [9] для анализа одномерных сигналов в нейросетях применяется ряд Фурье.

В работах [6–8] используются спектральные преобразования для улучшения СНС. Данные работы направлены на сохранение информации при уменьшении размерности. Однако, преобразование Уолша менее ресурсоемкое, чем вейвлет-преобразование и преобразование Фурье, и подходит для обработки изображений с БПЛА. Гибридная архитектура адаптирована под сравнение пар изображений, а не под классификацию/восстановление.

В работе [10] исследуется вейвлет анализ для обработки изображений с БПЛА, включая задачи навигации. Авторы комбинируют СНС с частотным разложением для улучшения выделения признаков. В [10] применяется бинарная классификация (обнаружение объектов). В отличие от [10], где вейвлеты применяются для предобработки, наш метод интегрирует преобразование Уолша непосредственно в обучаемую архитектуру, что повышает адаптивность к вариациям входных данных.

В работе [11] решается задача сопоставления изображений, снятых в схожих условиях. Предложенная архитектура нейронной сети работает с парой изображений, полученных с разных сенсоров, что критически важно для автономной навигации в условиях отсутствия GPS. Работа [12] посвящена обзорам проблем навигации, которые необходимо учитывать в задачах автонавигации. В работе [13] рассматривается двухуровневая система управления. В исследованиях [11, 14] используют стандартную архитектуру и классическими методами подвыборки.

Визуальный поиск. Определение визуального сходства – это задача, которая выявляет насколько идентичны два изображения, используя их визуальное содержание. Дано m изображений из поисковой базы данных $G = \{I_1, I_2, \dots, I_M\}$. Каждое изображение $I_j, j = \overline{1, M}$ характеризуется вектором числовых значений признаков $\Phi(I) = \langle \varphi_1(I), \varphi_2(I), \dots, \varphi_N(I) \rangle$, где N – число признаков (мощность пространства признаков) R^N . Определим $\tau_{k,s} = \tau(I_k, I_s)$ как показатель сходства, который измеряет насколько подобны два изображения $I_k \in G$ и $I_s \in G$. Решаемая задача состоит в изучении функций визуального сходства τ , которые вычисляют значение показателя сходства, основываясь на векторах числовых значений признаков, как:

$$\tau_{k,s} \geq \tau_{min}, \quad (1)$$

где τ_{min} – пороговое значение сходства.

Решение задачи определения визуального сходства состоит из следующих этапов:

1. Формирование вектора числовых значений признаков изображения, которые определяют наиболее заметную визуальную информацию;
2. Численная оценка визуального сходства (мера близости) на основе сформированных векторов [15, 16].

Для реализации указанных этапов применим сиамскую архитектуру нейронной сети. Так как в качестве входных значений необходимо подавать пару изображений, то разработанную архитектуру расширим до сиамской. Выходным значением, является численная оценка сходства, чем она меньше, тем изображения визуально ближе. В нашем случае рассматривается географическая близость. Географическая близость двух изображений, находящихся в одной локации, но полученные с разных сенсоров должны $\rightarrow 0$.

Сиамская архитектура. Сиамская архитектура нейронной сети – это особый тип нейронных сетей, который предназначен для сравнение пары входных данных и определения степени их схожести. Данная архитектура применяется в задачах, где необходимо измерить сходство или расстояние между объектами [17, 18].

К ключевым особенностям данной архитектуры можно отнести:

1. Две идентичные подсети («близнецы»):
 - ◆ обе ветви используют одинаковые веса;
 - ◆ обучаются параллельно на парах входных данных.
2. Функции сравнения:
 - ◆ контрастная (Contrastive Loss);
 - ◆ triplet Loss;
 - ◆ косинусное расстояние:

$$\text{cosine}_{\text{similar}}(A, B) = \frac{(A, B)}{\|A\| \cdot \|B\|}, \quad (2)$$

где (A, B) – скалярное произведение векторов A и B , $\|A\|$ – норма (длина) вектора A , вычисляемая как $\sqrt{\sum_{i=1}^M A_i^2}$, i – индекс суммирования, M – размер каждого из векторов A и B , $\|B\|$ – норма (длина) вектора B , вычисляемая аналогично.

- ◆ MSE (Mean Squared Error) – среднеквадратичная ошибка.

3. Выходное значение СНС – степень схожести: число от 0 до 1, где 0 – это идентичные объекты.

На рис. 1 приведена модель вычисления сходства с использованием сиамских сетей [19].

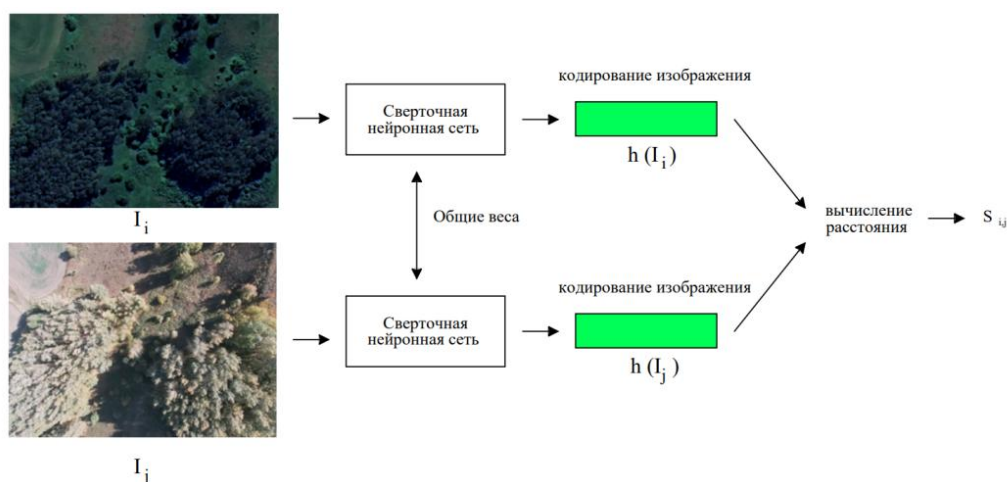


Рис. 1. Модель вычисления сходства с использованием сиамских сетей

Операции для уменьшения размерности. В данной работе рассматривались следующие операции для уменьшения размерности:

1. **Линейная интерполяция** – альтернативный подход к подвыборке, при котором используется взвешенная линейная интерполяция значений к некоторой точке. Модель на основе обучения выбирает к какой точке ближе аппроксимировать.

Алгоритм работы линейной интерполяции с обучаемым параметром.

1. Разделить входной тензор на блоки 2×2 .
2. Вычислить взвешенные коэффициенты для каждого пикселя в блоке, чтобы приблизить значение, например, к точке x_{11} .
3. Задаем смещение, как обучаемый параметр, что позволяет модели автоматически выбирать, к какой точке внутри блока аппроксимировать.

4. Для каждой точки в блоке задаем обучаемые веса, которые определяют их вклад в распределении.
5. Сформировать выходной тензор меньшего размера.
6. Передать выходной тензор дальше по нейронной сети.

Данный подход уменьшения размерности является альтернативой AvgPooling, но применяется тогда, когда важна именно плавность изменений, например, при спектральных преобразованиях. Данный подход может быть применен к нейронным сетям с малым количеством данных, так как в них меньше риск к переобучению. К преимуществам использования можно отнести то, что данная операция меньше искажает границы объектов, чем AvgPooling. Веса адаптивны к данным.

2. **Max Pooling** – классический метод подвыборки, который осуществляет уменьшение размерности карт признаков.
3. **WalsPooling** – это особый тип операции подвыборки, основанный на преобразовании Уолша. В отличие от стандартных методов (MaxPooling, AvgPooling), которые берут максимальное или среднее значение из региона, WalsPooling использует преобразование Уолша для агрегации признаков, что позволяет:
 - ♦ улучшить сжатие признаков с учетом их спектральных свойств;
 - ♦ сохранить больше информации, чем при стандартной подвыборке.

Алгоритм работы WalsPooling.

1. Применить прямое преобразование.

Данные получены после сверточного слоя. Размер передаваемой матрицы 256x256. Применяем прямое преобразование Уолша. Данная операция выполняется по следующей формуле:

$$\mathbf{W} \in \{-1, +1\}^{256 \times 256}$$

$$\mathbf{Y} = \mathbf{W} \cdot \mathbf{X} \cdot \mathbf{W}^T, \quad (3)$$

где $\mathbf{W}$ – матрица Уолша, $\mathbf{X}$ – входной сигнал.

2. Фильтрация низкочастотных составляющих.

Применим фильтр низких частот, обнулив 75% значений спектральной составляющей (рис. 2,б,д).

Чтобы их обнулить, применим маску $\mathbf{M}$:

$$M_{ij} = \begin{cases} 0, & \text{если } i + j \leq K \\ 1, & \text{иначе} \end{cases}, \quad (4)$$

где K – порог обнуления.

Результирующая спектральная матрица

$$\mathbf{W}' = \mathbf{W} \cdot \mathbf{M}, \quad (5)$$

где $\mathbf{M}$ – поэлементное умножение.

3. Применить обратное преобразование Уолша.

$$\mathbf{X}' = \frac{1}{256^2} \mathbf{W} \cdot \mathbf{W}' \cdot \mathbf{W}. \quad (6)$$

На рис. 2 приведена пара сравниваемых изображений, полученная с разных сенсоров: а – изображение с БПЛА, г – тайл подстилающей поверхности, б, д – фильтр низких частот, в, е – изображение после обратного преобразования.

4. Выполнить субдискретизацию.

После прямого преобразования и фильтрации производится обратное преобразование и субдискретизация (downsampling), выбирается только один элемент из каждого блока 2x2. На рис. 3 приведена демонстрация субдискретизации в качестве операции подвыборки.

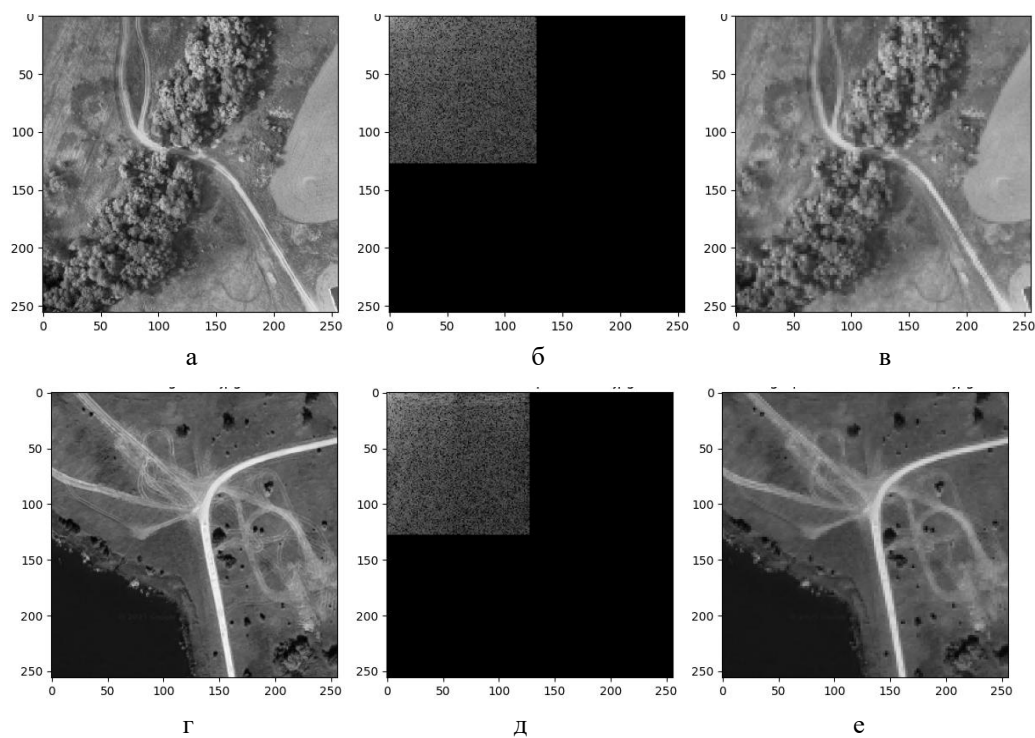


Рис. 2. Пара изображений после обратного преобразования Уолша

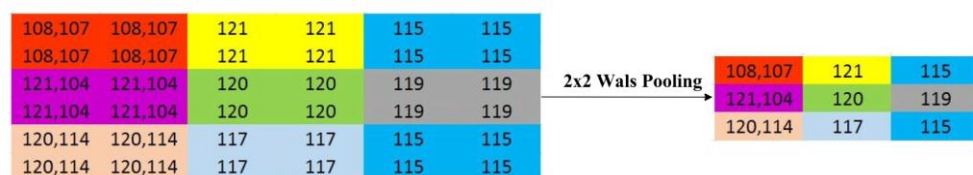


Рис. 3. Демонстрация операции подвыборки WalsPooling

5. Передать данные дальше по нейронной сети.

На рис. 4 приведена визуализация первого и второго слоя нейронной сети с различными операциями для уменьшения размерности. Для простоты восприятия в качестве входного изображения используем черно-белое изображение. На рис. 5 приведены гистограммы распределения значений, полученных после третьей свертки с применением преобразования Уолша для ядра свертки.

Преимуществом использования WalsPooling является быстрое вычисление, исходя из того, что матрица Уолша состоит только из -1 и +1.

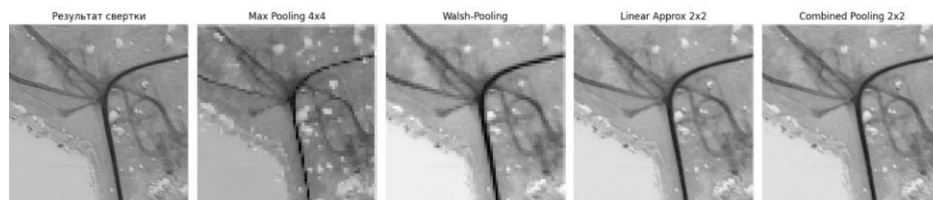


Рис. 4. Визуализация первого и второго слоя спектральной нейронной сети с различными операциями для уменьшения размерности

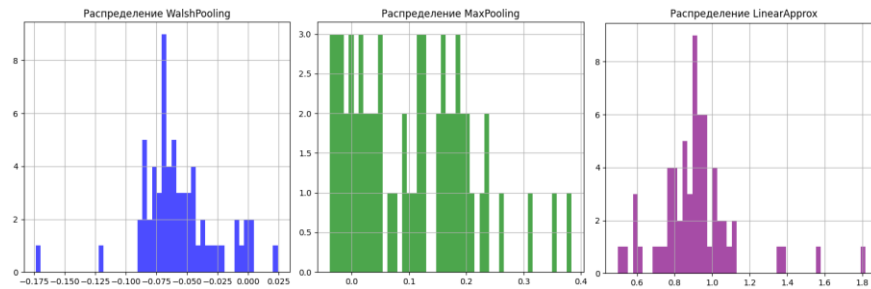


Рис. 5. Гистограммы распределения значений для различных операций уменьшения размерности

Исходя из рис. 5 видно, что гистограмма распределения значений для операции WalshPooling и для линейной аппроксимации к центру квадрата идентична. Ниже приведено сравнение характеристик выходов:

Сравнение характеристик выходов:
 WalshPooling: min=0.7067, max=1.5253, mean=0.9676
 MaxPooling: min=0.9746, max=3.2710, mean=1.7445
 LinearApprox: min=0.4994, max=1.8163, mean=0.9187

Финальные предсказания:
 WalshPooling: 0.650653
 MaxPooling: 0.558279
 LinearApprox: 0.279787

Однако, при схожей гистограмме распределений операций WalshPooling и LinearApprox финальные предсказания отличаются. Финальное предсказание – это выходы нейросети после прохождения всех слоев, включая пулинг и полносвязные слои, сжатые до одного числа в диапазоне $[0, 1]$ (благодаря сигмоиде в последнем слое). WalshPooling и LinearApprox имеют одинаковые min, max и mean (гистограммы близки), но финальные предсказания у них сильно различаются (0.6506, 0.2089).

- ◆ WalshPooling $\rightarrow$ 0.6506.
- ◆ MaxPooling $\rightarrow$ 0.5583.
- ◆ LinearApprox $\rightarrow$ 0.279787.

Это означает, что

◆ WalshPooling даёт предсказания ближе к 1, что необходимо в задачах классификации.

- ◆ LinearApprox сильно занижает выход (возможно, из-за сглаживания).
- ◆ MaxPooling – промежуточный вариант, но ближе к WalshPooling.

Для того, чтобы улучшить результат, можно рассмотреть:

- ◆ комбинацию методов (WalshPooling + MaxPooling);
- ◆ изменить матрицу Уолша (например, использовать только низкочастотные компоненты).

Архитектура гибридной свёрточной нейронной сети. На вход нейронной сети подаются два изображения, полученные с разных сенсоров. На основе визуального сходства осуществляется сравнение пары изображений. Выходной слой свёрточной нейронной сети дает численную оценку близости пары изображений. В качестве операций для уменьшения размерности в слоях нейронной сети используется линейная интерполяция, подвыборка Max Pooling и разработанный WalsPooling. WalsPooling – это особый тип операции подвыборки, основанный на преобразовании Уолша. Рассмотрены гистограммы распределения значений после операций подвыборки.

На вход свёрточной нейронной сети подаются изображения, после предварительной обработки, уменьшенное до 256x256. В качестве предварительной обработки применим спектральное преобразование Уолша к изображению, затем удалим низкочастотную составляющую и восстановим изображение [20–22].

На рис. 6 приведена модель предлагаемой сиамской нейронной сети.

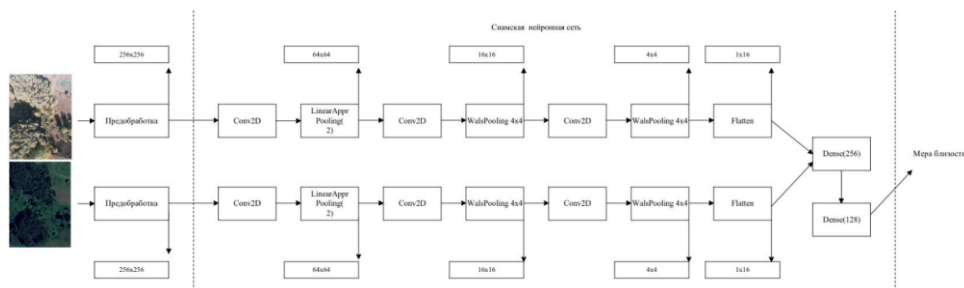


Рис. 6. Модель предлагаемой сиамской нейронной сети

Рассмотрим **разработанную архитектуру** свёрточной нейронной сети:

1. Исходное изображение 256x256 преобразуется в спектр Уолша, удаляются низкочастотные компоненты и восстанавливается изображение.
2. Применяем операция свертки к предобработанному изображению с ядром свёртки, вычисленным на основе обучения нейронной сети.
3. К полученной после операции свёртки 256x256 применяем линейную интерполяцию. Тем самым получаем 64x64 значения операции свёртки после усеечения.
4. Применяем операцию свёртки 64x64 для полученной на шаге 3 сверки (после усеечения).
5. К полученной после операции свёртки 64x64 применяем операцию усеечения (WalsPooling 4x4), с помощью слоя нейронной сети, представленного в частотной области. Получим 16x16 значения операции свёртки после усеечения.
6. Применяем к полученной после усеечения WalsPooling операцию свёртки с ядром свёртки, вычисленным на основе обучения нейронной сети.
7. Применяем к полученной свертке 16x16 операцию усеечения (WalsPooling 4x4). Получим 4x4 значений.
8. Полученные 4x4 значений преобразуем из многомерного пространства в одномерный вектор, чтобы его можно было подать на вход полносвязному слою (Dense) в нейронной сети. Получим 16 значений.
9. Следующие два полносвязных слоя осуществляют линейное преобразование + активацию. На первом полносвязном слое на вход поступает 16 нейронов с предыдущего слоя и 256 весов.

$$h_9 = \sigma(W_9 x_{flat} + b_9), \quad (7)$$

где x_{flat} – вектор, полученный на 8 шаге, $W_9 \in R^{256 \times 16}$ – матрица весов, $b_9 \in R^{256}$, $\sigma(z)$ – сигмоидная активация. Итого на выходе с данного слоя получим 4096+256=4352 параметров.

10. Второй полносвязный слой Dense(128). На вход поступает 256 нейронов с предыдущего слоя. Итого на выходе 32768+128=32896 параметров.

Dense(256) принимает 16 значений и выдаёт 256 значений → 4352 параметра. Dense(128) принимает 256 значений и выдаёт 128 значений → 32,896 параметров. Общее число параметров до сиамской части: 37,248.

11. После шагов 9 и 10 формируется две ветви (для изображений A и B):

$$x_{11}^a, x_{11}^b \in R^{128},$$

12. Общие слои для обеих ветвей. Повторим два полносвязных слоя Dense(256), Dense(128).

13. Сравнение ветвей осуществляется с использованием вычитания. Применяется квадратичная активация.

14. Финальный слой Dense(1). На вход подается вектор из 128 значения и 1 смещения. Получаем 129 параметров, которые описывают степень соответствия.

Полные формулы предложенной архитектуры

Для двух изображений A и B:

1. Спектральные свертки и подвыборка:

$$\begin{aligned} x_{flat}^a &= Flatten(WalshPooling(Conv3(Walsh(A)))) \\ x_{flat}^b &= Flatten(WalshPooling(Conv3(Walsh(B)))) \end{aligned} \quad (8)$$

2. Общие полносвязные слои:

$$\begin{aligned} x_{11}^a &= \sigma(W_{10}(\sigma(W_9 x_{flat}^a + b_9) + b_{10})) \\ x_{11}^b &= \sigma(W_{10}(\sigma(W_9 x_{flat}^b + b_9) + b_{10})). \end{aligned} \quad (9)$$

3. Сиамская сеть:

$$y = w_{out} [\sigma(W_{14}\sigma(W_{13}x_{11}^a + b_{13}) + b_{14}) - \sigma(W_{14}\sigma(W_{13}x_{11}^b + b_{13}) + b_{14})]^2 + b_{out}. \quad (10)$$

На рис. 5,6 приведена визуализация работы предложенной архитектуры гибридной свёрточной нейронной сети. Различия зависят от ядра свёртки, которое формируется в процессе обучения нейронной сети.

Визуализация слоев нейронной сети. На рис. 7 и 8 приведена визуализация слоев для пары изображений, полученных с разных сенсоров. На рис. 7, 8:

- а) – результат после первой операции свёртки;
- б) – результат после операции уменьшения размерности (линейная аппроксимация к центру квадрата);
- в) – результат после второй операции свёртки;
- г) – результат после подвыборки разработанным WalshPooling;
- д) – результат после операции свёртки;
- е) – результат после подвыборки разработанным WalshPooling.

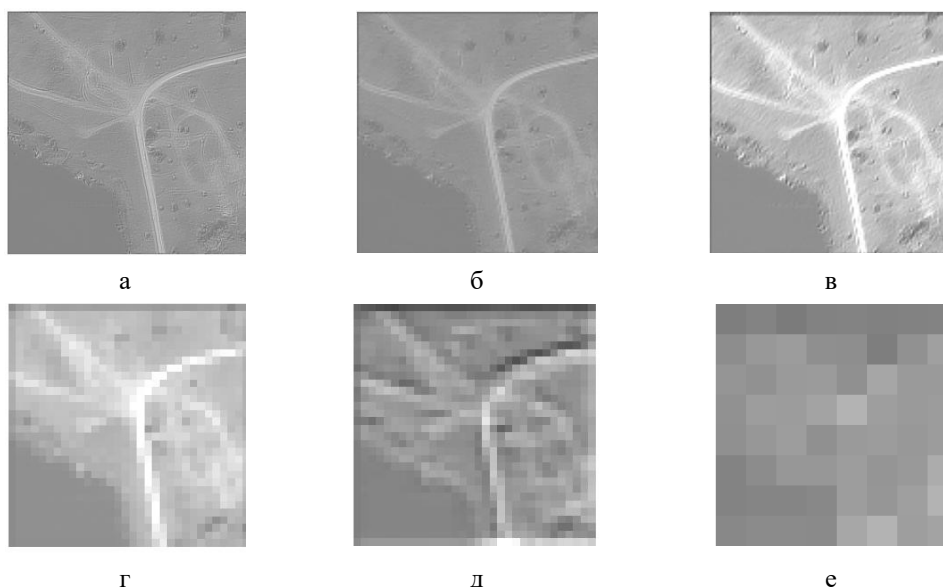


Рис. 7. Визуализация работы 6 слоев гибридной свёрточной нейронной сети для снимка с БПЛА

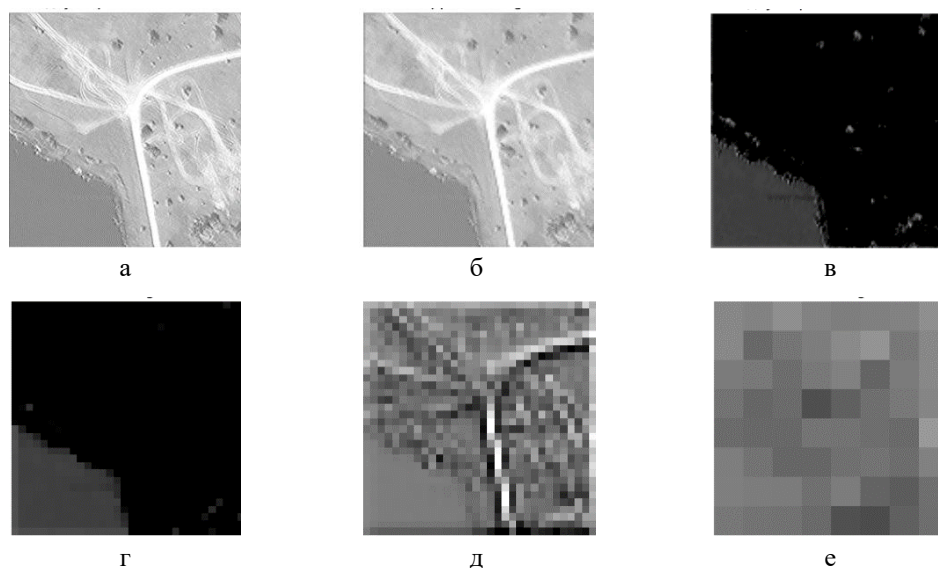


Рис. 8. Визуализация работы 6 слоев гибридной свёрточной нейронной сети для тайла подстилающей поверхности

В данной архитектуре отсутствуют активационные функции, что позволяет снижать размерность слоев нейронной сети, а также ускоряет сборку и обучение нейронной сети. Однако, можно добавить в разработанную архитектуру активационные функции и сравнить сходимости сети с ними и без них.

Подход к формированию набора данных для обучения и результаты экспериментов. Проведем эксперимент на разработанной архитектуре. В данной статье предложен подход к формированию данных для обучения нейронной сети. В качестве искажений применяются следующие виды трансформаций (табл. 2).

Таблица 2

Виды трансформации

Категория трансформации	Тип трансформации	Базовое значение и разброс
Яркостные искажения	Изменение яркости	$1.0 \pm 0.3\%$
	Изменение контрастности	$1.0 \pm 0.3\%$
	Добавление Гауссова шума	0.0 ± 0.2
Геометрические искажения	Поворот изображения	$0.0 \pm 5.0^\circ$
	Масштабирование	1.5 ± 0.25

Сравнение результатов работы сходимости нейронной сети производилось на:

- ◆ парах изображений без добавления шума и аугментации;
- ◆ парах изображений с добавлением шума и аугментации;
- ◆ парах изображений после применения спектрального преобразования Уолша и восстановленного изображения;
- ◆ парах изображений после применения спектрального преобразования Уолша и восстановленного изображения без учета постоянной составляющей;
- ◆ парах изображений после применения спектрального преобразования Уолша, усеченных на 25%, 50%, 75% значений;
- ◆ парах изображений восстановленного изображения, усеченных на 25%, 50%, 75% значений;
- ◆ парах изображений после спектральных преобразований с предварительным добавлением шума, аугментации и поворота;

В табл. 3 приведены результат работы сиамской нейронной сети.

Обозначения:

1. Без искажений.
2. Яркостные искажения.
3. Яркостные и геометрические искажения.
- a) Без удаления информационной составляющей.
- b) С удалением 25%.
- c) С удалением 50%.
- d) С удалением 75%.

x – Validation Accuracy (пары спутниковых снимков в разных локациях с искажениями).

y – Validation Loss (сред. квадр. ошибка).

z – Test Accuracy (пары спутник/беспилотник в идентичных локациях без искажений).

Таблица 3

Результат работы сиамской нейронной сети с разным набором входных значений

Тип входных данных	Спектр Уолша			Спектр Уолша без постоянной составляющей			Восстановленное изображение			Восстановленное изображение без постоянной составляющей		
	x	y	z	x	y	z	x	y	z	x	y	z
Метрика												
A-1	0.4975	2.4463	1	0.5013	2.2463	1	0.7494	1.1650	7	0.7400	1.0977	5
A-2	0.5088	2.6302	1	0.4937	2.4682	1	0.7800	1.0922	6	0.7325	1.0530	5
A-3	0.5075	2.5378	1	0.4900	3.0640	1	0.6875	1.4485	4	0.6784	1.5660	5
B-1	0.5300	2.3623	2	0.4586	2.3105	1	0.8175	0.7654	5	0.7600	1.2870	5
B-2	0.5000	2.6127	1	0.4727	2.4101	1	0.7675	0.5376	7	0.7800	0.8082	6
B-3	0.5125	2.5174	1	0.4782	2.4473	0	0.7750	0.7742	6	0.7118	1.1856	7
C-1							0.7085	0.8285	6	0.7775	1.2558	4
C-2							0.7425	0.9872	7	0.7450	1.1599	6
C-3							0.7250	0.8841	5	0.7425	1.1982	7
D-1							0.7500	1.1539	8	0.7025	1.1021	4
D-2							0.7965	0.9484	6	0.7100	1.2198	6
D-3							0.7450	0.8754	4	0.7425	1.2830	6

Низкий показатель **x** – Validation Accuracy (пары спутниковых снимков в разных локациях с искажениями). **y** – Validation Loss (сред. квадр. ошибка) на наборах данных, где в качестве входного изображения выступает спектральное представление связано с тем, что сверточная нейронная сеть в операциях свёртки оперирует не частотным представлением, поэтому необходимо переходить от пространственного представления к частотному. Наилучший результат СНС показывает в том случае, если в качестве входного слоя применяется восстановленное изображение, причем можно убирать до 75% значений, в том числе и постоянную составляющую. Лучшая конфигурация: Стандартные свёртки с восстановлением дают максимальную Validation Accuracy (0.7965) и приемлемый Loss (0.9484).

Метрики информационного поиска. Для того, чтоб оценить точность поиска рассмотрим следующие метрики:

♦ точность поиска (precision) вычисляется как отношение выданных релевантных изображений к общему количеству выданных изображений;

$$p = \frac{RF}{RF+IF}, \quad (11)$$

где RF – количество выданных релевантных изображений, IF – количество выданных нерелевантных изображений.

♦ полнота поиска (recall) вычисляется как отношение выданных релевантных изображений к общему числу релевантных изображений;

$$r = \frac{RF}{RF+RN}, \quad (12)$$

где RN – количество релевантных не выданных значений.

F1-мера (сбалансированная F-мера, мера Ван Ризбергена, F1 – measure) применяется как единая мера, объединяющая полноту и точность, которая вычисляется следующим образом:

$$F_1 = \frac{2rp}{r+p}. \quad (13)$$

F1-мера (значение F1-меры не может быть ниже полноты и точности и превышать их среднеарифметическое):

$$\min(r, p) \leq F_1 \leq \frac{r+p}{2}; \quad (14)$$

Для предложенной архитектуры получим следующие результаты, приведенные в табл. 4, 5.

Таблица 4

Метрики информационного поиска для предложенной архитектуры

Метрика	Значение	Интерпретация
Precision	0.728	72.8% пар, которые модель назвала "похожими", действительно похожи
Recall	0.800	Модель находит лишь 80.0% всех реально похожих пар
F1	0.872	Значение 0.872 близко к 1, что указывает на высокое качество модели в балансе между: precision (точностью) и recall (полнотой).
AUC-ROC	0.784	Значение 0.784 означает, что в 78.4% случаев модель верно ранжирует случайно выбранный положительный пример выше, чем отрицательный.

Таблица 5

Средние расстояния между парами

Класс	Значение	Интерпретация
Same (похожие)	0,317	Среднее расстояние для пар одной локации
Diff (разные)	1.869	Среднее расстояние для пар разных локаций

Заключение. В работе предложена гибридная архитектура СНС, сочетающая спектральные и пространственные слои для оценки географической близости изображений, полученных с разных сенсоров: спутниковое изображение и изображение, полученное с БПЛА. Введены новые методы подвыборки и свёртки, демонстрирующие устойчивость к изменениям масштаба и угла съёмки. Использована сямская архитектура для обработки пар изображений, что критично для задач сопоставления. В дальнейшем предлагается расширить набор данных для обучения нейронной сети с разными условиями съёмки (погода, время суток) для покрытия большего числа сценариев, ввести дополнительные метрики IoU (Intersection over Union) для оценки геометрического перекрытия предсказанных и истинных областей.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Gaihua Wang, Guoliang Yuan, Meng Lv, WenZhou Liu. Application of Interpolation Pooling in Convolutional Neural Networks // Helix. – 2018. – Vol. 8 (4). – P. 3465- 3469. – DOI 10.29042/2018-3465-3469
2. Zhuo Chen, Jiyuan Zhang, Ruizhou Ding. ViP: Virtual Pooling for Accelerating CNN-based Image Classification and Object Detection // WACV. – 2020 – P. 1180-1189. – https://openaccess.thecvf.com/content_WACV_2020/papers/Chen_ViP_Virtual_Pooling_for_Accelerating_CNN-based_Image_Classification_and_Object_WACV_2020_paper.pdf.

3. *Travis Williams, Robert Li.* Wavelet Pooling for Convolutional Neural Networks // Published as a conference paper at ICLR. – 2018. – P. 1-12. – <https://openreview.net/forum?id=rkhlb8lCZ>.
4. *Juan Manuel Fortuna-Cervantes, Marco Tulio Ramírez-Torres, Marcela Mejía-Carlos, José Salomé Murguía, José Martínez-Carranza, Carlos Soubervielle-Montalvo, César Arturo Guerra-García.* Texture and Materials Image Classification Based on Wavelet Pooling Layer in CNN // the Special Issue Recent Advances in Deep Learning for Image Analysis. – <https://doi.org/10.3390/app12073592>.
5. Image-Net. – <https://www.image-net.org/>.
6. *Pengju Liu, Hongzhi Zhang, Wei Lian, Wangmeng Zuo.* Multi-Level Wavelet Convolutional Neural Networks // Digital Object Identifier. – 10.1109/ACCESS.2019.2921451. – <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8732332>.
7. *Вершков Н.А., Бабенко М.Г., Черных А.Н., Кучуков В.А., Кучеров Н.Н., Кучукова Н.Н.* Оптимизация искусственных нейронных сетей с помощью вейвлет-преобразований // Программирование. – 2022. – № 6. – С. 22-30. – DOI: 10.31857/S0132347422060073.
8. *Spitsyn V.G., Bolotova Yu.A., Phan N.H., Bui T.T.T.* Using a Haar wavelet transform, principal component analysis and neural networks for OCR in the presence of impulse noise. – <https://computeroptics.ru/eng/KO/Annot/KO40-2/400216.html>.
9. *Митина О.А., Базанова Д.В.* Разработка модели нейронной сети на основе ряда Фурье // НАУ. – 2023. – № 90-2. – URL: <https://cyberleninka.ru/article/n/razrabotka-modeli-neyronnoy-seti-na-osnove-ryada-furie> (дата обращения: 12.08.2025).
10. *Фортуна-Сервантес Х.М., Рамирес-Торрес М.Т., Мартинес-Карранса Х., Мургуиа-Ибарра Х.С., Мехиа-Карлос М.* Обнаружение объектов в аэронавигации с использованием вейвлет-преобразования и сверточных нейронных сетей: первый подход // Тр. ИСП РАН. – 2021. – Т. 33. – Вып. 2. – DOI: 10.15514/ISPRAS-2020-33(2)-9.
11. *Беляев П.Ю., Зикратов И.А.* Исследование автономной навигации беспилотных летательных аппаратов на основе корреляционных методов сравнения изображений // Тр. учебных заведений связи. – 2024. – Т. 10, № 5. – С. 109-118.
12. *Давыдова А.А., Чечиков Ю.Б.* Проблема реализации современной GPS-навигации для БПЛА // Перспективные научные исследования: опыт, проблемы и перспективы развития: Сб. научных статей по материалам X Международной научно-практической конференции. – Уфа, 2023. – С. 100-106.
13. *Кориков А.М., Тран В.Т.* Применение нейронных сетей и корреляционно-экстремального координатора для навигации и управления БПЛА // Природные и интеллектуальные ресурсы Сибири (СИБРЕСУРС-28-2022): Доклады (материалы конференции) 28-й международной научно-практической конференции. – Томск, 2022. – С. 29-33.
14. *Лесив Е.А., Вытовтов К.А., Барабанова Е.А.* Нейросетевой алгоритм анализа видеoinформации для резервной системы локальной навигации привязных БПЛА с обучением на синтетических данных // Информационные технологии и технические средства управления (ICCT- 2023). Матер. VII Международной научной конференции. – М., 2023. – С. 159-162.
15. *Ustinova V.E., Lutsenko A.S., Shpak A.V., Mironenkov G.V., Ivlev V.A.* A method for finding the correspondence between a railway station model and its visual representation based on graphs // Computing, Telecommunications and Control. – 2024. – Vol. 17, No. 4. – P. 64-77.
16. *Ustinova V.E., Lutsenko A.S., Shpak A.V., Mironenkov G.V., Ivlev V.A.* A method for finding the correspondence between a railway station model and its visual representation based on graphs // Computing, Telecommunications and Control. – 2024. – Vol. 17, No. 4. – P. 64-77.
17. *Кочетуров В.В., Федяев О.И.* Сиамские свёрточные нейронные сети для распознавания лиц // Программная инженерия: методы и технологии разработки информационно-вычислительных систем (ПИИВС-2024): Сб. материалов и докладов V Международной научно-практической конференции. – Донецк, 2024. – С. 178-184.
18. *Ладанова Е.О.* Разработка методики и новой структуры сиамской нейронной сети для анализа геопространственных данных // Научно-технический вестник Поволжья. – 2024. – № 12. – С. 144-147.
19. *Sumina E.A.* Siamese networks as a possible improvement of standard algorithms // Artificial Intelligence: Challenges, Essence, Communication, Creativity and other aspects: Сб. материалов студенческих конференций. – М., 2024. – С. 104-108.
20. *Куртяник Д. В., Сергеев А.М., Самиков А.В., Семенов А.А.* О визуальной оценке результатов поиска матриц Адамара // Наука. Техника. Технологии (политехнический вестник). – 2018. – № 4. – С. 19-26.
21. *Кушков С.И., Цветков А.В.* Основы теории сигналов: учеб. пособие. – Екатеринбург: Изд-во Урал. ун-та, 2023. – 114 с. – ISBN 978-5-7996-3636-4.
22. *Костров Б.В., Гринченко Н.Н., Баранова С.Н., Трушина Е.А., Вьюгина А.А.* Ортогональное кодирование бинарных изображений // Вестник Ярославского высшего военного училища противовоздушной обороны. – 2023. – № 2 (21). – С. 82-87.

REFERENCES

1. Gaihua Wang, Guoliang Yuan, Meng Lv, WenZhou Liu. Application of Interpolation Pooling in Convolutional Neural Networks, *Helix*, 2018, Vol. 8 (4), pp. 3465- 3469. DOI 10.29042/2018-3465-3469
2. Zhuo Chen, Jiuyan Zhang, Ruizhou Ding. ViP. Virtual Pooling for Accelerating CNN-based Image Classification and Object Detection, *WACV*, 2020, pp. 1180-1189. Available at: https://openaccess.thecvf.com/content_WACV_2020/papers/Chen_ViP_Virtual_Pooling_for_Accelerating_CNN-based_Image_Classification_and_Object_Detection_WACV_2020_paper.pdf.
3. Travis Williams, Robert Li. Wavelet Pooling for Convolutional Neural Networks, *Published as a conference paper at ICLR*, 2018, pp. 1-12. Available at: <https://openreview.net/forum?id=rkhlb8ICZ>.
4. Juan Manuel Fortuna-Cervantes, Marco Tulio Ramirez-Torres, Marcela Mejia-Carlos, José Salomé Murguía, José Martínez-Carranza, Carlos Soubervielle-Montalvo, César Arturo Guerra-García. Texture and Materials Image Classification Based on Wavelet Pooling Layer in CNN, *the Special Issue Recent Advances in Deep Learning for Image Analysis*. Available at: <https://doi.org/10.3390/app12073592>.
5. Image-Net. Available at: <https://www.image-net.org/>.
6. Pengju Liu, Hongzhi Zhang, Wei Lian, Wangmeng Zuo. Multi-Level Wavelet Convolutional Neural Networks, *Digital Object Identifier*. 10.1109/ACCESS.2019.2921451. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8732332>.
7. Vershkov N.A., Babenko M.G., Chernykh A.N., Kuchukov V.A., Kuchеров N.N., Kuchukova N.N. Optimizatsiya iskusstvennykh neyronnykh setey s pomoshch'yu veyvlet-preobrazovaniy [Optimization of artificial neural networks using wavelet transforms], *Programmirovaniye* [Programming], 2022, No. 6, pp. 22-30. DOI: 10.31857/S0132347422060073.
8. Spitsyn V.G., Bolotova Yu.A., Phan N.H., Bui T.T.T. Using a Haar wavelet transform, principal component analysis and neural networks for OCR in the presence of impulse noise. Available at: <https://computeroptics.ru/eng/KO/Annot/KO40-2/400216.html>.
9. Mitina O.A., Bazanova D.V. Razrabotka modeli neyronnoy seti na osnove ryada Fur'e [Development of a neural network model based on the Fourier series], *NAU* [National Association of Scientists], 2023, No. 90-2. Available at: <https://cyberleninka.ru/article/n/razrabotka-modeli-neyronnoy-seti-na-osnove-ryada-furie> (accessed 12 August 2025).
10. Fortuna-Servantes Kh.M., Ramires-Torres M.T., Martinez-Karransa Kh., Murguia-Ibarra Kh.S., Mekhia-Karlos M. Obnaruzhenie ob"ektov v aeronavigatsii s ispol'zovaniem veyvlet-preobrazovaniya i svertochnykh neyronnykh setey: pervyy podkhod [Object Detection in Air Navigation Using Wavelet Transform and Convolutional Neural Networks: A First Approach], *Tr. ISP RAN* [Proceedings of ISP RAS], 2021, Vol. 33, Issue 2. DOI: 10.15514/ISPRAS-2020-33(2)-9.
11. Belyaev P.Yu., Zikratov I.A. Issledovanie avtonomnoy navigatsii bespilotnykh letatel'nykh apparatov na osnove korrelyatsionnykh metodov sravneniya izobrazheniy [Study of Autonomous Navigation of Unmanned Aerial Vehicles Based on Correlation Methods of Image Comparison], *Tr. uchebnykh zavedeniy svyazi* [Proceedings of Communications Educational Institutions], 2024, Vol. 10, No. 5, pp. 109-118.
12. Davydova A.A., Chechikov Yu.B. Problema realizatsii sovremennoy GPS-navigatsii dlya BPLA [The problem of implementing modern GPS navigation for UAVs], *Perspektivnye nauchnye issledovaniya: opyt, problemy i perspektivy razvitiya: Sb. nauchnykh statey po materialam X Mezhdunarodnoy nauchno-prakticheskoy konferentsii* [Advanced scientific research: experience, problems and development prospects: Collection of scientific articles based on the materials of the X International scientific and practical conference]. Ufa, 2023, pp. 100-106.
13. Korikov A.M., Tran V.T. Primenenie neyronnykh setey i korrelyatsionno-ekstremalnogo koordinatora dlya navigatsii i upravleniya BPLA [Application of neural networks and correlation-extreme coordinator for navigation and control of UAVs], *Prirodnye i intellektual'nye resursy Sibiri (SIBRESURS-28-2022): Doklady (materialy konferentsii) 28-y mezhdunarodnoy nauchno-prakticheskoy konferentsii* [Natural and intellectual resources of Siberia (SIBRESURS-28-2022): Reports (conference materials) of the 28th international scientific and practical conference]. Tomsk, 2022, pp. 29-33.
14. Lesiv E.A., Vytovtov K.A., Barabanova E.A. Neyrosetevoy algoritim analiza videoinformatsii dlya rezervnoy sistemy lokal'noy navigatsii privyaznykh BPLA s obucheniem na sinteticheskikh dannykh [Neural network algorithm for analyzing video information for a backup local navigation system for tethered UAVs with training on synthetic data], *Informatsionnye tekhnologii i tekhnicheskie sredstva upravleniya (ICCT- 2023). Mater. VII Mezhdunarodnoy nauchnoy konferentsii* [Information technology and technical means of control (ICCT-2023). Proceedings of the VII International Scientific Conference]. Moscow, 2023, pp. 159-162.
15. Ustinova V.E., Lutsenko A.S., Shpak A.V., Mironenkov G.V., Ivlev V.A. A method for finding the correspondence between a railway station model and its visual representation based on graphs, *Computing, Telecommunications and Control*, 2024, Vol. 17, No. 4, pp. 64-77.

16. Ustinova V.E., Lutsenko A.S., Shpak A.V., Mironenkov G.V., Ivlev V.A. A method for finding the correspondence between a railway station model and its visual representation based on graphs, *Computing, Telecommunications and Control*, 2024, Vol. 17, No. 4, pp. 64-77.
17. Kocheturov V.V., Fedyaev O.I. Siamskie svertochnye neyronnye seti dlya raspoznavaniya lits [Siamese convolutional neural networks for face recognition], *Programmnaya inzheneriya: metody i tekhnologii razrabotki informatsionno-vychislitel'nykh sistem (PIIVS-2024): Sb. materialov i dokladov V Mezhdunarodnoy nauchno-prakticheskoy konferentsii* [Software Engineering: Methods and Technologies for Developing Information and Computing Systems (PIIVS-2024): Collection of materials and reports of the V International scientific and practical conference]. Donetsk, 2024, pp. 178-184.
18. Ladanova E.O. Razrabotka metodiki i novoy struktury siamskoy neyronnoy seti dlya analiza geoprostranstvennykh dannyykh [Development of a methodology and a new structure of a Siamese neural network for analyzing geospatial data], *Nauchno-tekhnicheskii vestnik Povolzh'ya* [Scientific and Technical Bulletin of the Volga Region], 2024, No. 12, pp. 144-147.
19. Sumina E.A. Siamese networks as a possible improvement of standard algorithms, *Artificial Intelligence: Challenges, Essence, Communication, Creativity and other aspects: Collection of materials from student conferences*. Moscow, 2024, pp. 104-108.
20. Kurtyanik D. V., Sergeev A.M., Samikov A.V., Semenov A.A. O vizual'noy otsenke rezul'tatov poiska matrits Adamara [On the visual assessment of the results of the search for Hadamard matrices], *Nauka. Tekhnika. Tekhnologii (politekhnicheskii vestnik)* [Science. Engineering. Technologies (Polytechnic Bulletin)], 2018, No. 4, pp. 19-26.
21. Kumkov S.I., TSvetkov A.V. Osnovy teorii signalov: ucheb. posobie [Fundamentals of signal theory: a tutorial]. Ekaterinburg: Izd-vo Ural. un-ta, 2023, 114 p. ISBN 978-5-7996-3636-4.
22. Kostrov B.V., Grinchenko N.N., Baranova S.N., Trushina E.A., V'yugina A.A. Ortogonal'noe kodirovaniye binarnyykh izobrazheniy [Orthogonal coding of binary images], *Vestnik Yaroslavskogo vysshego voennogo uchilishcha protivovozdushnoy oborony* [Bulletin of the Yaroslavl Higher Military School of Air Defense], 2023, No. 2 (21), pp. 82-87.

Костров Борис Васильевич – Рязанский государственный радиотехнический университет; e-mail: kostrov.b.v@evm.rsreu.ru; г. Рязань, Россия; тел.: +79036930570; профессор; зав. кафедрой электронных вычислительных машин.

Бабаев Сергей Игоревич – Рязанский государственный радиотехнический университет; e-mail: babaev.s.i@gmail.com; г. Рязань, Россия; тел.: +79209676677; доцент кафедры электронных вычислительных машин.

Ефимов Алексей Игоревич – Рязанский государственный радиотехнический университет; e-mail: lexie62rus@mail.ru; г. Рязань, Россия; тел.: +79105665083; доцент кафедры электронных вычислительных машин.

Тарасова Валентина Юрьевна – Рязанский государственный радиотехнический университет; e-mail: Valentina2008.91@mail.ru; тел.: +79038373769; старший преподаватель кафедры электронных вычислительных машин.

Kostrov Boris Vasilyevich – Ryazan State Radio Engineering University; e-mail: kostrov.b.v@evm.rsreu.ru; Ryazan, Russia; phone: +79036930570; professor, head of Department of Electronic Computers.

Babaev Sergey Igorevich – Ryazan State Radio Engineering University; e-mail: babaev.s.i@gmail.com; Ryazan, Russia; phone: +79209676677; associate Professor, Department of Electronic Computers.

Efimov Alexey Igorevich – Ryazan State Radio Engineering University; e-mail: lexie62rus@mail.ru; Ryazan, Russia; phone: +79105665083; associate professor, Department of Electronic Computers.

Tarasova Valentina Yurievna – Ryazan State Radio Engineering University; e-mail: Valentina2008.91@mail.ru; Ryazan, Russia; phone: +79038373769; senior Lecturer, Department of Electronic Computers Machines.

В.Б. Савинов, Н.Н. Шушарина**ОБЗОР МЕТОДОВ УЛУЧШЕНИЯ РАССУЖДЕНИЙ В БОЛЬШИХ
ЯЗЫКОВЫХ МОДЕЛЯХ**

Появление больших языковых моделей стало важным этапом в области обработки естественного языка, поскольку такие модели демонстрируют впечатляющие результаты в генерации, трансформации и анализе текстов, а также в решении широкого спектра прикладных задач. Однако, несмотря на значительные успехи в практическом применении, большие языковые модели обладают ограниченными способностями к рассуждению. Эти ограничения проявляются в трудностях обобщения знаний за пределами обучающего распределения, проблемах переноса знаний в новые контексты, а также в снижении точности выполнения многошаговых логических и математических операций. Целью данной работы является изучение методов повышения способности больших языковых моделей к рассуждению, где под рассуждением понимается процесс формирования и оценки выводов из существующей информации. В работе рассматриваются основные типы рассуждений, релевантные для больших языковых моделей: математическое, логическое и рассуждение на основе здравого смысла. Приводится список наиболее часто используемых контрольных тестов, применяемых для оценки качества рассуждений языковых моделей. Проводится обзор методов улучшения рассуждений, применяемых в больших языковых моделях на 2025 год. В зависимости от этапа применения (на этапе обучения или на этапе использования языковой модели) - рассматриваются подходы к подготовке данных для обучения, изменения в архитектуре языковых моделей, процедуры обучения и дообучения моделей (в том числе с использованием специально подготовленных синтетических наборов данных), обучение с подкреплением, различные варианты построения смысловых цепочек, механизмы интеграции внешних инструментов и многоагентный подход. Также рассматриваются существующие ограничения больших языковых моделей, которые проявляются в отсутствии понимания концепций, плохой обобщающей способности вне распределения обучающих данных и снижении эффективности при росте сложности задач. Выделяются наиболее перспективные методы, направленные на повышение качества и надежности рассуждений больших языковых моделей.

Большая языковая модель; методы рассуждений; искусственный интеллект.

V.B. Savinov, N.N. Shusharina**A REVIEW OF METHODS FOR IMPROVING REASONING IN LARGE
LANGUAGE MODELS**

The emergence of large language models has become an important milestone in the field of natural language processing, as such models demonstrate impressive results in text generation, transformation, and analysis, as well as in solving a wide range of applied tasks. However, despite significant practical success, large language models possess limited reasoning capabilities. These limitations manifest in difficulties with generalizing knowledge beyond the training distribution, challenges in transferring knowledge to new contexts, and reduced accuracy when performing multi-step logical and mathematical operations. The goal of this work is to examine methods for improving the reasoning abilities of large language models, where reasoning is understood as the process of forming and evaluating inferences based on existing information. The paper discusses the main types of reasoning relevant to large language models: mathematical, logical, and commonsense reasoning. It provides a list of the most commonly used benchmarks applied to assess the reasoning quality of language models. An overview is presented of the methods used to enhance reasoning in large language models at 2025. Depending on the stage of application (during training or during model usage), the work examines approaches to training data preparation, architectural modifications of language models, training and finetuning procedures (including those using specially constructed synthetic datasets), reinforcement learning, various chain-of-thought construction techniques, mechanisms for integrating external tools, and multi-agent approaches. The paper also discusses existing limitations of large language models, which include the lack of conceptual understanding, poor out-of-distribution generalization, and reduced effectiveness as task complexity increases. Finally, the most promising methods aimed at improving the quality and reliability of reasoning in large language models are highlighted.

Large language model; reasoning methods; artificial intelligence.

Введение. Появление больших языковых моделей (БЯМ) вызвало настоящий прорыв в области обработки естественного языка (англ. Natural Language Processing, NLP). Лежащая в основе БЯМ архитектура трансформер демонстрирует впечатляющую производительность в прогнозировании следующего токена [1] и кроме задач, связанных с обработкой текста, успешно применяется в задачах компьютерного зрения, обработки мультимодальных данных, системах управления роботами и т.д. Однако БЯМ обладают ограниченными способностями к рассуждению, что обусловлено их конструкцией и процедурой обучения, которая делает акцент на статистическом моделировании языка, а не на понимании [2]. Это поднимает вопрос о возможностях БЯМ к проявлению способности к рассуждению, что является одним из важнейших аспектов систем искусственного интеллекта (ИИ).

Целью данной работы является изучение методов, применяемых для улучшения рассуждений БЯМ. В первой части описывается, что понимается под рассуждением в контексте БЯМ. Во второй части перечисляются контрольные тесты, используемые для оценки рассуждений. Третья часть рассматривает методы используемые для улучшения рассуждений БЯМ, исходя из возможных этапов применения при разработке БЯМ: подготовка данных, выбор архитектуры, обучение, дообучение и использование. Приводятся существующие у БЯМ ограничения к рассуждению. В заключении делаются выводы о наиболее перспективных методах для улучшения качества рассуждений БЯМ.

1. Рассуждение в больших языковых моделях. Большая языковая модель (от англ. large language model, LLM) – это предварительно обученная статистическая языковая модель (рассматривает текст как последовательность слов и выполняет оценку вероятности слов), основанная на искусственных нейронных сетях (ИНС) с большим числом параметров [3]. В основном, БЯМ построены на архитектуре трансформер [1] и содержат от единиц до сотен миллиардов параметров [3].

Критики БЯМ отмечают их способность совершать глупые ошибки (фактические и логические), непоследовательность и ограниченность рассуждений, отсутствие здравого смысла и планирования ответа, невозможность исправления их авторегрессионной сути. БЯМ демонстрируют неспособность обобщения информации [4] и существенное снижение точности ответов в случае нестандартного представления задач [5].

Однако области применения БЯМ регулярно расширяются и кроме задач обработки естественного языка они уже применяются для решения самых разных задач: решение математических проблем, управление роботами и автономными автомобилями, реализация различных автономных агентов [6], социологические исследования и т.д.

Рассуждение – это процесс формирования и оценки выводов из существующей информации [7]. Традиционно, рассуждения разделяются на дедуктивное, индуктивное и абдуктивное, но применительно к базовым моделям ИИ, рассуждение разделяют по задачам применения: рассуждения на основе здравого смысла, математические рассуждения, логические рассуждения, математические рассуждения, причинно-следственные рассуждения, визуальные рассуждения, аудио рассуждения, мультимодальные рассуждения, воплощенные рассуждения и т.д. В случае БЯМ, рассуждение разделяют на математическое, логическое, рассуждение на основе здравого смысла и рассуждение специфичное для предметной области [8].

Несмотря на все успехи БЯМ в генерации и обработке текста, требуется различать формальную и функциональную языковые компетенции [9]. Где под формальной языковой компетенцией понимается знание правил и статистических закономерностей языка, а под функциональной языковой компетенцией – способность к языковому мышлению. Учитывая это различие, можно отметить, что хотя БЯМ владеют формальной языковой компетенцией на уровне человека, с функциональной языковой компетенцией отмечают проблемы [9]. При этом, если с увеличением количества обучающих данных формальная языковая компетенция у БЯМ также улучшается, то с функциональной языковой компетенцией эффект масштаба выражен менее явно, что вынуждает разработчиков достигать интересных видов поведения БЯМ при помощи специализированных методов (обучение с подкреплением или использование внешних инструментов).

Таким образом, несмотря на существующие проблемы и критику – появляются новые области и способы применения БЯМ, а представления о возможностях БЯМ постоянно меняются. Применительно к БЯМ, основными видами рассуждения являются математическое, логическое и рассуждение на основе здравого смысла. Стандартные БЯМ обладают ограниченной способностью к рассуждению и хотя они способны решать задачи, похожие на образцы, присутствовавшие в обучающей выборке, но в случае новых проблем, требующих рассуждений - у БЯМ возникают трудности.

2. Тесты для оценки производительности. Для контроля прогресса развития БЯМ используются специальные контрольные тесты (бенчмарки) [10].

Для оценки рассуждений на основе здравого смысла используются:

- ◆ HellaSwag – выбор из нескольких вариантов завершения предложения.
- ◆ BoolQ – вопрос по короткому тексту из Википедии с вариантами ответа Да/Нет.
- ◆ PIQA – вопросы на физический здравый смысл.
- ◆ SIQA (SocialQA – Social Interaction QA) – описание социальных ситуаций и возможных вариантов ответов.

◆ WinoGrande – решение неоднозначных задач заполнения пропусков в предложениях, которые предлагают двоичные варианты (стал ответом на достижение высоких результатов на Winograd Schema Challenge).

◆ ARC (AI2 Reasoning Challenge) – вопросы по естественным наукам начальной школы (разделен на Сложный и Простой наборы вопросов).

Для оценки математических рассуждений БЯМ применяются:

- ◆ MATH – решение математических задач.
- ◆ GSM8K – тестирование математических рассуждений. Набор состоит из порядка 8000 вопросов и ответов по математике для начальной школы (7473 обучающихся и 1319 тестовых задач).
- ◆ MGSM (англ. Multilingual Grade School Math) – это 250 задач из GSM8K, переведенные на 10 языков (Испанский, Французский, Немецкий, Русский, Китайский, Японский, Тайский, Суахили, Бенгальский, Телугу).
- ◆ SVAMP (англ. Simple Variations on Arithmetic Math word Problems) – набор элементарных математических текстовых задач (состоят из короткого текстового описания и вопросов о неизвестных значениях).
- ◆ AIME (англ. American Invitational Mathematics Examination) – набор математических олимпиадных задач для американских школьников (например, AIME24 и AIME25 – 30 задач).

Для оценки логических рассуждений используются:

- ◆ ProntoQA – синтетический набор данных для анализа цепочки мыслей БЯМ.
- ◆ LogiQA – набор из 8678 вопросов для проверки понимания прочитанного и оценки способности к логическому мышлению (данные: 7376 для тренировки, 651 для оценки и 651 для тестирования) на Английском и Китайском языках).

Другие контрольные тесты, используемые для проверки БЯМ:

- ◆ GPQA (Graduate-Level Google-Proof Q&A Benchmark) – сложный набор из 448 вопросов (от экспертов по биологии, физики и химии) с несколькими вариантами ответов.
- ◆ BIG-bench (Beyond the Imitation Game benchmark) – набор для оценки способностей БЯМ. Состоит из 204 разнообразных заданий (включают: лингвистику, детское развитие, математику, здравый смысл, биологию, физик, социальные предубеждения, разработку программного обеспечения).
- ◆ BIG-Bench Hard (BBH) – набор из 23 сложных задач BIG-bench, требующих многоэтапных рассуждений (включают в себя логическую дедукцию, арифметику, рассуждения на основе здравого смысла).
- ◆ DROP (Discrete Reasoning Over the content of Paragraphs) – набор для оценки возможностей по пониманию и рассуждению. Набор содержит 96 тысяч вопросов, которые были получены с помощью краудсорсинга (открытки из Википедии и сложные вопросы к ним).

♦ AlpacaEval – метод для оценки способности следовать инструкциям пользователя. Набор данных содержит 805 пользовательских инструкций с эталонными ответами от GPT-4. Для уменьшения предвзятости, связанной с предпочтением более длинных ответов используется AlpacaEval 2.

♦ Arena-Hard – метод для оценки ответов на запросы пользователя. Набор данных постоянно обновляется на основе данных с краудсорсинговой платформы Chatbot Arena, содержит 500 пользовательских запросов и использует GPT-4-Turbo в качестве судьи для сравнения с ответами базовой модели (GPT-4).

♦ ARC-AGI (англ. Abstraction and Reasoning Corpus for Artificial General Intelligence) – тест на интеллект общего назначения, нацеленный на способности обобщения и адаптации к новым задачам (публичные данные: 400 тренировочных и 400 тестовых задач, приватные данные: 200 задач для валидации).

Для оценки способностей БЯМ к рассуждению предлагаются и более простые тесты: например, подсчет букв в слове (человек может выполнить эту задачу с точностью в 100%, а от БЯМ потребуется базовая способность к рассуждениям).

Однако при использовании контрольных тестов необходимо учитывать следующие проблемы: насыщение (на некоторых бенчмарках современные БЯМ достигают почти идеальной производительности), утечки тестовых данных (тестовые данные бенчмарков могли попасть в тренировочную выборку БЯМ, что может приводить к переобучению модели), качество самих тестов (бенчмарки различаются по качеству и могут иметь сложности с воспроизводимостью). Для преодоления этих сложностей появляются новые более сложные бенчмарки, разрабатываются динамические или закрытые версии контрольных тестов.

Таким образом, для тестирования БЯМ существует большое количество самых разнообразных бенчмарков. Для оценки способности БЯМ к рассуждению наиболее часто используются ARC AI2, GSM8K (или MGSM), MATH, BBH.

3. Методы улучшения рассуждений. Для улучшения способности БЯМ к рассуждениям возможно использовать различные подходы. Рассмотрим их исходя из возможных этапов применения: на этапах обучения (подготовка данных, архитектура, дообучение) или на этапе использования БЯМ (рис. 1).



Рис. 1. Методы улучшения рассуждений БЯМ.

Данные. Качественные и разнообразные данные играют важнейшую роль при обучении языковых моделей. При создании открытого набора данных FineWeb-Edu было показано, что фильтрация интернет-страниц из наборов данных CommonCrawl и вычленение из них только обучающих материалов – позволило повысить показатели в тестах оценки знаний и рассуждений [11]. Тщательная фильтрация и вычленение страниц с ма-

тематическими данными из набора CommonCrawl и данные программного кода позволили обучить модель DeepSeekMath 7B, которая продемонстрировала 51.7% на математическом бенчмарке MATH [12]. Включение в обучающую выборку данных с программным кодом повышает не только способность БЯМ генерировать код, но также улучшает способность к рассуждению и знания о мире [13].

Распространенным подходом улучшения рассуждений является обучение на синтетических данных.

Используя качественные образцы и принципы символической логики, были получены образцы логических рассуждений в виде набора данных Formal Logic Deduction Diverse (FLD) (состоит из примеров многоступенчатых дедукций с неизвестными фактами, разными правилами рассуждений), что позволило улучшить способность БЯМ к логическим рассуждениям [14].

В процессе обучения модели phi-4 (14 миллиардов параметров) в основном использовались синтетические данные, что позволило показать производительность на уровне моделей большего размера, особенно в контрольных тестах на рассуждение [15]. Для создания наборов синтетических данных в качестве основы можно использовать предварительно отобранные данные из сети Интернет, которые затем можно использовать для создания задач на рассуждение [15].

При обучении модели MiniMax-M1, отказались от использования синтетических данных, а отдали приоритет парам вопрос-ответ и до 70% увеличили долю данных, соответствующих науке, технологии, инженерии и математике (англ. Science, Technology, Engineering, Mathematics, STEM) [16].

Таким образом, использование при обучении БЯМ качественных данных позволяет значительно улучшить производительность моделей, без необходимости масштабирования их размеров. Важным методом является использование синтетических данных, но требуется внимательно контролировать их качество и разнообразие.

Архитектура. Рассмотрим какие изменения можно внести в архитектуру трансформера, чтобы добиться улучшения качества рассуждений.

В основе архитектуры трансформера лежит механизм внимания, использующий многопеременную логистическую функцию (softmax), которая применяется для расчета итоговых весов внимания, переводя итоговый вектор значений в вероятностное распределение. При работе обученной модели с данными, чье распределение отличается от того, которое использовалось при обучении, softmax размывается. Чтобы исправить ситуацию и добиться более резкого распределения, предлагается использовать адаптивную температуру [17]. Также механизм внимания может быть модифицирован для формирования паттернов ассоциации и манипуляции, что важно для выполнения рассуждений, включающих связывание и преобразование понятий [18].

Изменения в самой архитектуре БЯМ, включающие добавление обучаемого механизма связей между слоями, могут способствовать улучшению математических рассуждений [19].

Учитывая, что исследования нейронаук показывают, что человеческий мозг (как носитель обобщенного интеллекта) имеет разные области, отвечающие за обработку языка и решение логических и математических задач, то и для построения будущих интеллектуальных систем потребуется не единственная модель, а модульная архитектура, сочетающая языковые модели с моделями, которые представляют абстрактные знания и поддерживают сложные рассуждения [9].

Исследование показывает, что рассуждения можно проводить не выходя в пространство токенов [20]. У БЯМ может быть реализована работа в двух режимах: латентном – когда внутреннее представление БЯМ используется для последующей генерации, и языковом – стандартный режим работы БЯМ с генерацией токенов. Процесс обучения подобной модели начинается с обычной генерации текста, но на последующих этапах – первые шаги рассуждения заменяются внутренним представлением (выделяются специальными токенами <bot> и <eot>). Подобный подход позволяет модели выполнять процесс рассуждения без языковых ограничений, что может интерпретироваться в виде поиска по дереву [20].

Высказывается предположение, что для осуществления когнитивных процедур, связанных с обработкой информации (в частности – рассуждений) мозг опирается на пространственные представления (характеризующие не только мир вещей, но и субъективный внутренний мир мыслей), приобретаемые при взаимодействии с окружающим миром [21]. Для объединения БЯМ с моделями, осуществляющими обработку сенсорных данных или робототехнических данных - может использоваться подход смешения мнений экспертов (англ. Mixture of experts, MoE).

К архитектурным изменениям также относится изменения токенизатора (инструмент, преобразующий текст в данные (токены)). Подход «Байтовый латентный трансформер» (англ. Byte Latent Transformer, BLT) демонстрирует отказ от токенизатора с фиксированным словарем и переход на динамическую токенизацию уровне байтов (входной поток разбивается на наборы байтов, чьи границы определяются по энтропии следующего байта) [22]. Данный подход не только демонстрирует эффективность обучения и вывода за счет динамического выбора длины наборов байт (если данные предсказуемы), но и улучшение в рассуждениях и обобщении [22].

Таким образом, для улучшения способностей и качества рассуждений можно рассматривать различные варианты изменения архитектуры БЯМ: изменение токенизатора, механизма внимания, добавление возможности обработки сенсорной информации, использование архитектуры MoE.

Обучение. Для улучшения качества рассуждений, БЯМ обучаются в несколько этапов, которые характеризуются увеличением относительного количества высококачественных данных с примерами рассуждений (математические задачи, код) [23].

Так предварительное обучение моделей Qwen3 и MiMo-7B [23] проходило в три этапа:

1. Начальный этап обучения, на котором модель получает предварительные знания. Используется небольшое контекстное окно (для моделей Qwen3 и MiMo-7B: 4096 и 8192 токенов, соответственно).

2. Повышение способности к рассуждению, за счет увеличения доли качественных (научных, математических, программного кода, примеров рассуждения) данных. Длина контекстного окна не меняется.

3. Расширение возможностей решения сложных задач за счет дополнительного включения высококачественных данных, синтетических ответов на задачи по математике и программированию. Расширение длины контекста (для моделей Qwen3 и MiMo-7B: до 32768 токенов).

Для предварительного обучения модели MiniMax-M1 использовались 7.5T токенов данных: на первых 2.5T токенах данных обучение велось с постоянной скоростью обучения $8e-5$, а затем постепенно уменьшалась до $8e-6$ [16]. Расширение длины контекста происходило постепенно в 4 этапа (позволяет избежать взрыва градиентов), начиная от 32 килобайт и увеличивая до 1 миллиона токенов.

Таким образом, для улучшения у БЯМ способности к рассуждению, процедура предварительного обучения проводится в несколько этапов, с постепенным увеличением количества качественных данных, относящихся к математическим задачам, примерам программного кода, научным и техническим материалам.

Дообучение. Для улучшения рассуждений БЯМ можно использовать дообучение модели на специально подготовленных данных.

Метод самообучения рассуждению (англ. Self-Taught Reasoner, STaR) использует БЯМ, чтобы генерировать рассуждения для ответа на вопросы, используя несколько первоначальных примеров рассуждений в качестве образца [24]. В случае если сгенерированный ответ неверен - рассуждение генерируется снова (учитывая правильный ответ), а полученные рассуждения (давшие правильный ответ) используются в качестве данных для дообучения модели, и далее цикл повторяется. Данный подход позволил на основе небольшого набора первоначальных подсказок (с несколькими примерами) создать большой набор данных рассуждений, что позволило дообученной модели улучшить результаты на контрольных тестах CommonsenseQA и GSM8K [24]. Однако одной из клю-

чевых проблем подобного подхода является возможность использования неправильных обоснований, которые в конце приводят к правильным ответам. Подобные рассуждения будут попадать в обучающую выборку, что приведет к снижению качества модели. Для преодоления данной проблемы – необходимо проверять не только ответ, но и все промежуточные шаги рассуждения.

Для обучения БЯМ способности генерировать внутренние размышления перед формированием ответа была предложена методика оптимизации мыслительных предпочтений (англ. Thought Preference Optimization, TPO), которая заключается в итеративной генерации БЯМ различных вариантов размышлений и ответов, оценке их качества с помощью отдельной модели-судьи и последующей оптимизации БЯМ с помощью прямой оптимизации предпочтений (англ. Direct Preference Optimization, DPO) на основе этих предпочтений [25]. Модель (Llama-3-8B-Instruct), обученная с помощью TPO, продемонстрировала улучшение результатов не только в задачах, требующих рассуждения, но и в общих знаниях, что позволяет предположить универсальность подхода. Однако, хотя тесты на AlpacaEval 2 и Arena-Hard показали улучшение – результаты на GSM8K стали хуже, что может быть связано с отсутствием достаточного числа математических инструкций во время обучения [25].

Для решения задач ARC-AGI, используя комбинированный подход индукции (выведение промежуточных функций (программный код на Python)) и трансдукции (непосредственное предсказание результатов, используя только имеющиеся примеры), БЯМ обучили на большом корпусе (сотни тысяч) синтетических данных программного кода, генерирующего варианты решения задач, что позволило приблизиться к уровню производительности человека [26].

Метод самокоррекции с помощью обучения с подкреплением (англ. Self-Correction via Reinforcement Learning, SCoRe) улучшает способность БЯМ самостоятельно исправлять свои ошибки, используя многоэтапный процесс обучения с подкреплением, используя данные, сгенерированные самой моделью [27]. Метод SCoRe повысил точность базовой модели на бенчмарках MATH (+15.6%) и HumanEval (+9.1%) [27].

Метод обратного мышления (англ. Reverse Thinking, RevThink), вдохновляется человеческой способностью рассуждать как в прямом, так и в обратном направлениях и позволяет улучшить способность БЯМ к рассуждению путем дообучения модели (Mistral-7B-Instruct-v0.3, с помощью LoRA (англ. Low-Rank Adaptation)) на данных, представляющих собой сгенерированные решения сразу трех задач: прямое рассуждение по вопросу, формулировка обратного вопроса и обратные рассуждения в ответ на обратный вопрос [28]. Метод RevThink не только показал лучшие результаты на бенчмарках: GSM8K(60.9%), MATH (15.3%), ARC (78.5%), но и продемонстрировал способность эффективно обобщать знания на новые, ранее не встречавшиеся наборы данных [28].

Дообучение на образцах рассуждения методом контролируемой тонкой настройки (англ. Supervised fine-tuning, SFT) даже на небольшом количестве математических примеров позволяет научить БЯМ генерировать смысловые цепочки [29, 30].

В [29] из начального набора в 59 тысяч математических вопросов, был отобран небольшой набор данных s1K, содержащий всего 1000 примеров. Каждый пример состоит из исходного вопроса, примера рассуждения и решения (сгенерированных Google Gemini Flash Thinking API). Процесс отбора этой тысячи примеров основывался на принципах качества, сложности и разнообразия:

- ♦ игнорировались некачественные и плохо отформатированные примеры,
- ♦ выбирались вопросы, которые не могли решить базовые БЯМ (Qwen2.5-7B-Instruct и Qwen2.5-32B-Instruct) и требовали для решения длительного рассуждения,
- ♦ для классификации вопросов по предметным областям использовалась отдельная БЯМ (Claude 3.5 Sonnet).

На полученном наборе данных s1K была дообучена (SFT) модель Qwen2.5-32B-Instruct, в результате чего была получена модель s1-32B, которая продемонстрировала значительное улучшение производительности (по сравнению с базовой моделью) на контрольных тестах AIME24, MATH500 и GPQA Diamond. Также был представлен простой

метод, принудительного ограничения бюджета ответа, позволяющий контролировать объем вычислений (токенов размышления), используемых во время генерации [29]. Данный метод либо принудительно завершает процесс рассуждений модели путем принудительного добавления фразы завершения рассуждений («Final Answer:»), либо стимулирования более продолжительных рассуждений путем удаления токена завершения рассуждений и добавления фразы «Wait», если модель пытается преждевременно остановиться. Подобный подход позволяет способствовать коррекции рассуждений и ответов модели, что было продемонстрировано увеличением точности на AIME24 по мере выделения большего количества токенов на размышление [29].

В [30] из миллионов математических задач было отобрано для последующего дообучения всего 817 примеров. При этом, основная гипотеза исследования состояла в том, что БЯМ уже имеют предварительные знания, скрытые в пространстве параметров модели и достаточно небольшого количества примеров, демонстрирующих образцы решения проблем, чтобы активировать способность БЯМ к сложным рассуждениям. Для отбора данных использовались критерии уровня сложности, общности и разнообразия:

- ◆ приоритет сложным задачам, способствующим развитию сложных цепочек рассуждений (сначала исключались задачи, которые за несколько попыток могла решить Qwen2.5-Math-7B-Instruct и оставляли только те задачи, которые рассуждающие модели DeepSeek-R1 и DeepSeek-R1-Distill-Qwen32B могли решить с вероятностью ниже заданного порога);

- ◆ выбор задач, которые стимулируют новые подходы к рассуждениям;

- ◆ должны охватывать широкий спектр математических задач.

Кроме отбора самих задач, осуществлялся отбор решений, качество которых оценивалось по структурной организации, эффективной реализации и надежности ответа:

- ◆ решение должно иметь четкую структуру и форматирование, избегая излишней многословности в объяснении простых шагов;

- ◆ поступательное решение с четким изложением ключевых идей;

- ◆ процесс рассуждения должен содержать этапы проверки (промежуточных результатов, предположений, логической последовательности выводов).

В результате дообучения (SFT) модели Qwen2.5-32B-Instruct на полученном наборе данных была получена модель LIMO, которая продемонстрировала улучшение производительности не только на математических контрольных тестах (AIME24, MATH500 и AMC23 (American Mathematics Competition)), но и на задачах, относящихся к областям вне пределов распределения тренировочных данных (англ. Out-of-Distribution, OOD).

Таким образом, дообучение БЯМ на специально подготовленных данных позволяет обучить БЯМ генерировать смысловые цепочки и улучшить качество рассуждений.

Обучение с подкреплением. Для выполнения сложных рассуждений, модель o1 от компания OpenAI была обучена с помощью обучения с подкреплением (англ. Reinforcement Learning, RL) [31]. Так как технические подробности о процедуре обучения модели o1 отсутствуют – возможно только рассмотреть возможные подходы по воспроизведению возможностей модели OpenAI o1 применительно к обучению с подкреплением [32]. Ключевыми компонентами являются: инициализация политики (предварительное обучение модели для развития навыков рассуждения), дизайн вознаграждения (создание эффективных сигналов для формирования вознаграждений), поиск (система генерации качественных решений на этапах обучения и тестирования) и само обучение (используя данные, полученные в процессе поиска, для улучшения политики) [32].

Появление открытой модели DeepSeek-R1 [33] продемонстрировало как именно обучение с подкреплением позволяет создавать большие рассуждающие модели (БРМ) (англ. large reasoning model, LRM). В основе DeepSeek-R1 лежит MoE-модель DeepSeek-V3 (671 миллиардов параметров), на базе которой была обучена модель DeepSeek-R1-Zero. При этом, использовалось только обучение с подкреплением, где в качестве алгоритма применялся подход оптимизации групповой относительной политики [12] (англ. Group Relative Policy Optimization, GRPO). Главной особенностью данного алгоритма является отказ от использования отдельной оценочной модели и выполнение оценки по группе сгенерированных ответов:

1. Формирование группы ответов на запрос.
2. Определение оценки для каждого ответа в группе.
3. Вычисление преимущества ответов относительно среднего значения оценок в группе.

Ответы с положительным преимуществом усиливают генерацию.

Оценка ответов модели осуществлялась при помощи проверок правильности ответа, корректности форматирования и выполнения проверочного кода.

Примечательно, что во время обучения у модели случился то, что авторы назвали «момент озарения» («Aha Moment»), когда в цепочке рассуждений модель написала: “Wait, wait. Wait. That’s an aha moment I can flag here.” и пересмотрела первоначальный подход к решению задачи [33].

Однако цепочки рассуждений DeepSeek-R1-Zero обладали ограниченной читабельностью, поэтому итоговая модель DeepSeek-R1 была обучена в несколько этапов:

1. Начальное дообучение базовой модели DeepSeek-V3-Base на специально подготовленном наборе данных, который получил название «Холодный старт» (англ. Cold Start). Данные были подготовлены в формате удобном для чтения человеком в виде: «блок рассуждения+итоговый ответ». Данный этап позволил дать более качественную начальную точку для последующего обучения.

2. Этап обучения с подкреплением, аналогичный DeepSeek-R1-Zero. В ответах модели наблюдалось смешение языков, поэтому было дополнительно добавлено вознаграждение за языковую согласованность (приводит к небольшому снижению производительности, но делает ответ более читаемым).

3. Дообучение (SFT) модели DeepSeek-V3-Base в течении 2 на специально подготовленном синтетическом наборе данных: 600 тысяч на рассуждение и 200 тысяч без рассуждений (ответы на вопросы, перевод).

4. Финальное обучение с подкреплением, аналогичное методологии обучения DeepSeek-R1-Zero, но в этом случае уже не только на математических и программных задачах, но и на общих запросах, где для оценки используются модели-оценщики, оценивающие полезность и безвредность ответа.

Итоговая модель DeepSeek-R1 не только продемонстрировала производительность на уровне передовой модели OpenAI o1, но и генерирует открытые результаты рассуждения.

Обучение с подкреплением превосходит контролируемую тонкую настройку в переносимости на другие задачи: модели, обученные с помощью RL на математических задачах также демонстрируют улучшение в задачах из других областей [34]. Наряду с RLHF, обучение с подкреплением и проверяемыми вознаграждениями (англ. Reinforcement Learning with Verifiable Rewards, RLVR) стали использоваться для дообучения моделей решению сложных задач. И если в случае RLHF используются данные о предпочтениях людей, то в RLVR – правильные результаты математических задач и выполнения инструкций (например, выполнения тестов для кода), что ограничивает возможности применения данного подхода в областях, где подобная проверка невозможна. Для преодоления данной проблемы, существует возможность использования собственной уверенности модели в качестве вознаграждения.

Таким образом, использование обучения с подкреплением позволяет значительно улучшить качество рассуждений БЯМ, однако его использование нуждается в дополнительном изучении.

Использование. БЯМ обучаются предсказывать токены (лексемы) и, фактически, они не предназначены для явных рассуждений, но используя определенные пользовательские инструкции можно решить эту проблему, направляя модель на выполнение необходимых шагов рассуждения [3]. К таким инструкциям относятся: цепочка и дерево мыслей, самосогласованность, рефлексия.

Цепочка мыслей (англ. Chain of Thought, CoT) является одним из самых распространенных методов улучшить способность БЯМ к рассуждению, используя только пользовательскую инструкцию [35]. Подход смысловой цепочки основывается на том, чтобы неявный процесс рассуждения стал явным: в инструкции описываются шаги рассуждения, необходимые чтобы БЯМ смогла прийти к логическому и обоснованному результату [3].

Существуют две основных формы подсказок цепочек мыслей (CoT):

1. С примерами (англ. Few-shot). В инструкции предоставляются несколько примеров пошаговых рассуждений, которые используются в качестве шаблонов для БЯМ [35].

2. Без примеров (англ. Zero-shot). В инструкцию добавляется фраза «Давайте размышлять шаг за шагом» («Let's think step by step»), что побуждает БЯМ самостоятельно формулировать этапы рассуждения [36].

Форма инструкции с примерами показывает большую эффективность, чем инструкция без примеров, но ее эффективность зависит от выбора подходящих и разнообразных примеров рассуждений [3]. Ручное конструирование подобных примеров является сложной и подверженной ошибкам задачей, что может быть решено автоматической CoT.

Дерево мыслей (англ. Tree of Thought, ToT) – это метод, основанный на идее рассмотрения альтернативных решений, представляя их в виде «дерева рассуждений», в котором каждая ветвь является отдельной линией рассуждений [37]. Данный метод позволяет исследовать различные возможные гипотезы, оценивая их обоснованность и релевантность исходной задаче [3]. ToT является расширением CoT и может быть полезен для решения сложных задач, где одной линии рассуждений может быть недостаточно (неоднозначность, учет нюансов).

Граф мыслей (англ. Graph of Thoughts, GoT) – метод моделирования информации, генерируемой БЯМ в виде графа, где вершины – это информационные единицы («мысли»), а ребра – зависимости между ними [38]. GoT является развитием таких подходов как CoT и ToT: при решении сложной задачи человек не только следует цепочке мыслей (как в CoT) или перебирает разные мысли (как в ToT), а формирует сложную сеть мыслей по которым может следовать, возвращаться и объединять между собой [38].

Самосогласованность (англ. Self-Consistency) – метод, где БЯМ генерирует несколько вариантов ответа (используя CoT) и согласованность этих ответов между собой, служит признаком правильности результата [39]. Данный метод основан на предположении, что в задачах, требующих рассуждения, возможно несколько различных путей решения, приводящих к одному правильному ответу. Самосогласованность не требует дополнительных данных, обучения, вспомогательных моделей и работает используя только одну БЯМ, но при этом повышает точность языковых моделей на контрольных тестах, требующих рассуждения (GSM8K, SVAMP, ARC, CommonsenseQA) [39]. Метод можно использовать для проверки фактов с целью обеспечения точности информации [3].

Самостоятельное уточнение (англ. Self-Refine) – метод улучшения первоначального ответа БЯМ посредством итеративной обратной связи и самокоррекции [40]. Особенность (и основное ограничение) метода состоит в том, что одна модель выполняет все функции: сначала генерирует первоначальный ответ, затем его оценивает (предоставляет обратную связь) и потом использует этот отзыв для улучшения своего ответа.

Подобный метод может быть реализован в виде самопроверки (англ. SelfCheck), когда БЯМ изучает и оценивает логические шаги (сгенерированные на предыдущих этапах), сравнивает результаты и исправляет ошибки [41].

Рефлексия – метод в котором БЯМ используется для оценки собственного ответа [42]. Метод является итеративным: после генерации первоначального ответа, модели предлагается оценить свой ответ, что позволяет модели выявить потенциальные ошибки (предложить улучшения) и уточнить свои результаты, повышая качество своих ответов [3].

Генерация ответа, дополненная результатами поиска (англ. Retrieval-augmented generation, RAG [43]) – метод сочетания внутренних знаний БЯМ с внешней информацией. Метод RAG-Star использует генерацию ответа, дополненную результатами поиска для улучшения процесса рассуждения [44]. В основе метода лежит использование поиска по дереву Монте-Карло (англ. Monte Carlo Tree Search, MCTS), используемого для итеративного планирования промежуточных подзапросов и ответов, необходимых для решения задачи (используются внутренние знания БЯМ), а дальнейший поиск по дереву выполняется с использованием RAG, чтобы оценивать рассуждения используя внешние данные (статьи из Википедии) [44].

Адаптация инструкций. Для решения сложных задач, обычной цепочки мыслей (CoT) может оказаться недостаточно. Метод SELF-DISCOVER предлагает подход по объединению различных техник рассуждений, когда БЯМ самостоятельно выбирает варианты, подходящие для конкретной задачи [45].

Сам метод состоит из трех шагов [45]:

1. Выбор. БЯМ выбирает несколько (наиболее подходящих для решения задачи) вариантов рассуждений из 39 шаблонов рассуждений (разбиение на подзадачи, креативное, критическое и системное мышление, рассуждение шаг за шагом и другие).

2. Адаптация. БЯМ предлагается адаптировать (перефразировать) выбранные варианты рассуждений для решения данной задачи.

3. Реализация. БЯМ использует адаптированную схему рассуждений для генерации пошагового плана рассуждений для получения финального ответа.

Подход SELF-DISCOVER показал улучшение результатов на контрольных тестах BIG-Bench Hard, Thinking for Doing и MATH, по сравнению с обычным CoT [45].

Обучение во время тестирования (англ. test-time training, TTT) – это расширение классического подхода (предварительное обучение модели и ее последующее использование без изменения параметров) на адаптацию модели для решения конкретной задачи [46].

Метод TTT отличается от обычного дообучения модели, малым количеством данных доступных для обучения, поэтому ключевыми компонентами успешного TTT являются: предварительное дообучение модели на схожих задачах, использование вспомогательных форматов задач и их аугментаций, обучение на каждом экземпляре отдельно (для обучения используется LoRA (для ускорения могут использоваться квантованные адаптеры – QLoRA)) и использование самосогласованности [46]. Использование TTT продемонстрировало значительное повышение точности на задачах ARC-AGI (улучшение в 6 раз по сравнению с базовыми моделями) [46].

Вычисления во время тестирования (англ. test-time compute) – это подход, суть которого состоит в увеличении объемов вычислений, используемых для вывода [47]. Исходя из предпосылки, что для решения сложных задач люди думают дольше, а БЯМ, используя дополнительные вычисления улучшают свои результаты в задачах, требующих рассуждения [35, 42] – было показано, что оптимальное распределение вычислительных ресурсов на этапе вывода может быть более эффективным, чем простое увеличение размера модели (особенно в задачах, требующих сложного рассуждения) [47].

Наиболее яркой демонстрацией успешности данного подхода стало появление осенью 2024 года модели GPT-o1 от компании OpenAI [31]. Модель GPT-o1 была обучена (с помощью обучения с подкреплением) на выполнение сложных рассуждений. Основным нововведением, которое отличает GPT-o1 от прошлых версий GPT стало явное включение цепочки рассуждений (CoT) во время выполнения процесса вывода, что позволяет модели создавать внутреннюю цепочку рассуждений и тем самым решать более сложные задачи [48]. Во время работы, модель выполняет несколько итераций генерации ответа в которых используются токены рассуждений (на следующей итерации генерации они отбрасываются, а в контекст добавляется только ответ, полученный на предыдущем шаге рассуждений).

Производительность модели o1 последовательно увеличивалась не только во время обучения (англ. train-time compute), но и при увеличении времени, затраченного на «размышления» модели (вычисления во время тестирования) [31]. В качестве контрольных тестов (вместо MATH и GSM8K) использовались результаты по AIME (олимпиада по математике для американских школьников) и GPQA. На всех бенчмарках (требующих рассуждения) модель o1 значительно превзошла GPT-4o [31].

Впечатляющие результаты, демонстрируемые моделью o1 от OpenAI, вдохновили и другие исследовательские команды на работы в направлении использования CoT и увеличения вычислений во время тестирования. Появились модели DeepSeek-R1-Lite-Preview и QwQ-32B-Preview.

Исследователи из Alibaba создали модель Marco-o1, в которой используется дообучение на цепочку мыслей, поиск по дереву Монте-Карло (MCTS) и механизм рефлексии [49]. В качестве базовой модели Marco-o1 используется Qwen2-7B-Instruct. Для дообуче-

ния модели (по всем параметрам) использовались наборы данных с примерами цепочек размышлений (Open-O1 CoT dataset (открытый), Marco-o1 CoT dataset (синтетический)). Для проверки модели использовались английский и китайский варианты из набора данных MGSM.

Исследователи HuggingFace также отметили тенденцию, когда вместо масштабирования вычислений во время обучения моделей — используется динамические стратегии вывода, которая позволяет моделям «дольше думать» [50]. Опираясь на исследования, проведенные DeepMind [47], которые показывают, что рост вычислений во время тестирования можно увеличивать для итеративного самосовершенствования или использования отдельной модели, оценивающей ответ для последующего поиска в пространстве решений — был рассмотрен вариант генерации большего числа решений и их последующая оценка отдельной моделью-оценщиком [50]. В качестве основной модели использовались Llama-3.2-Instruct (версии с 1 и 3 миллиардами параметров), а для оценки использовалась модель Llama3.1-8B-PRM-Deepseek-Data (модель с 8 миллиардами параметров, которая была специально обучена выдавать обратную связь на каждый этап процесса рассуждений). Первая модель генерирует решения (до 256 вариантов), а вторая — выполняет их оценку. Для оценки использовался набор данных MATH-500, который является подмножеством из набора данных MATH.

По результатам проведенного исследования, делаются выводы о важной роли правильной оценки шагов рассуждений (которые, как показывает контрольный тест ProcessBench, в настоящее время ограничены [51]), важности развития самопроверки [27], возможности улучшить рассуждения и принятие решений за счет включения промежуточных шагов рассуждений в процесс генерации, использование методов поиска для генерации обучающих наборов данных, адаптация данных методов на другие области рассуждений [50].

Развивая успех o1, OpenAI выпустила модель o3 (в конце декабря 2024), которая набрала 75.7% на оценочном наборе ARC-AGI, а высокопроизводительная конфигурация o3 набрала 87,5% (модели были предварительно обучены на тренировочном наборе), что означает значительное повышение уровня обобщения и адаптации.

Таким образом, среди существующих различных методов, позволяющих улучшить качество рассуждений у БЯМ, можно выделить следующие: подготовка качественных данных для обучения модели (включая программный код и синтетические данные), внесение изменений в архитектуру трансформера (например, использование MoE), обучение с подкреплением, дообучение на данных, содержащих различные варианты размышлений, и применение специальных методов использования БЯМ (CoT, ToT, GoT, самосогласованность, рефлексия, RAG, адаптация инструкций, обучение модели во время тестирования при помощи LoRA). Особо необходимо отметить подход вычисления во время тестирования, который демонстрирует наиболее впечатляющие результаты.

Применение внешних инструментов. БЯМ может использовать внешние инструменты, получая к ним доступ через вызовы программного интерфейса (англ. Application Programming Interface, API), что позволяет преодолевать присущие БЯМ ограничения в выполнении арифметических операций или получения актуальной информации из поисковых систем [52]. Также внешние инструменты могут использоваться для улучшения качества рассуждений.

Цепочка кода (англ. Chain of Code, CoC) — метод, предлагающий БЯМ использовать код для структурирования своих решений [53]. Ключевая идея CoC состоит в использовании БЯМ для представления рассуждений в виде подзадач (представленных программным псевдокодом), которые далее обрабатываются программным интерпретатором, а в случае неопределенного поведения интерпретатора — псевдокод обрабатывается БЯМ (выступающей в роли эмулятора) [53]. На бенчмарке BVH, CoC достигает 84%, что на 12% больше, чем у CoT (превосходит результаты людей в 18 из 23 задач) [53].

Программа мыслей (англ. Program of Thoughts, PoT) — метод в котором БЯМ генерирует программу (на языке программирования Python), которая описывает процесс рассуждения и далее выполняется внешним интерпретатором [54]. Сравнение PoT и CoT

показало рост на математических бенчмарках: GSM8K (+8.5% (71.6%)), SVAMP (+8.8% (85.2%)) [54]. Сочетание PoT с методом самосогласования позволяет добиться еще более лучших результатов: GSM8K (80.0%), SVAMP (89.1%), но они сильно зависят от используемой модели (приведенные результаты получены на модели OpenAI Codex) и зависят от примеров, использованных в инструкции БЯМ [54].

Аналогичный подход используется в *методе программно-управляемых языковых моделей* (англ. ProgramAided Language models, PAL), где БЯМ используется для преобразования задач, сформулированных на естественном языке, в программный код, который затем выполняется с помощью интерпретатора Python [55]. На математических бенчмарках PAL показало улучшение результатов, по сравнению с CoT: GSM8K (+6.4% (72.0%)), SVAMP (+4.6% (79.4%)) [55].

Таким образом, интеграция языковых моделей с внешними инструментами позволяет эффективно сочетать способности БЯМ в понимании и интерпретации задач с возможностями специализированных инструментов.

Агенты. Автономные агенты (АА) (система, взаимодействующая с внешней средой, способная воспринимать эту среду и действовать в ней, выполняя заданные задачи) – являются важным направлением исследований [6]. Системы на базе АА являются многообещающим подходом к созданию **искусственного интеллекта общего назначения** (англ. Artificial General Intelligence, AGI), который сможет выполнять задачи посредством самостоятельного планирования и действий.

Учитывая, что АА необходимо выполнять определенную задачу (роль), воспринимать окружающую среду и взаимодействовать с ней – их структура формализуется до модулей: профиля, памяти, планирования и действий [6]. Модуль профиля (составляется вручную или генерируется) определяет роль агента (программист, исследователь, редактор и т.п.), который записывается в инструкцию поступающую на вход БЯМ. Модуль памяти позволяет АА хранить информацию, полученную из окружающей среды и использовать эти данные для выполнения будущих действий. Модуль планирования позволяет АА планировать будущие действия (используя смысловые цепочки [35] или другие виды рассуждения), учитывая обратную связь от среды [56], человека или других моделей [40]. Модуль действий отвечает за взаимодействие с окружающей средой (преобразование решений агента в конкретные данные) посредством вызова внешних API, взаимодействия с интерпретаторами программного кода [54], базами данных или знаний и другими моделями.

Кроме различных примеров использования АА для социального моделирования, проведения экспериментов и различных видов автоматизации [6] – подход АА применяется и для улучшения рассуждений.

Метод рассуждения и действия (англ. Reasoning and Acting, ReAct) — метод предлагает БЯМ генерировать последовательность чередующихся рассуждений и действий (взаимодействие с внешними источниками для получения дополнительной информации), что позволяет модели контролировать и корректировать свой план действий [56].

Метод инструментально-интегрированных агентов рассуждений (англ. Tool-integrated Reasoning Agents, ToRA) применяет несколько агентов (использующих внешние инструменты (вычислительные библиотеки и символьные системы решения) для решения математических задач [57]. Для обучения ToRA использовался набор данных ToRA-Corpus (16 тысяч инструкций использования инструментов, сгенерированных при помощи GPT-4 на математических наборах данных GSM8K и MATH) [57]. В результате дообучения LLaMA-2 70B на ToRA-Corpus, полученная модель достигла следующих результатов: GSM8K (84.3%), MATH (49.7%), SVAMP (82.7%) [57].

Система мультиагентного обучения БЯМ (англ. Multiagent LLM training, MALT) предлагает структуру из трех специализированных агентов (генератор, проверяющий и улучшатель), которые последовательно взаимодействуют для решения поставленной задачи [58]. Базовая модель (Llama 3.1 8B) дообучается на синтетических данных траекторий рассуждения, используя систему распределения вознаграждений, основанную на совместном результате, что позволяет эффективно обучать каждую модель в системе и достичь следующих результатов: GSM8K (90.3%), MATH (56.5%) [58].

При этом, систематические эксперименты показывают, что одиночные БЯМ с качественными инструкциями способны достигать такого же уровня производительности, как и многоагентные системы, но многоагентные системы обсуждения превосходят одиночные модели в случаях, когда в инструкциях отсутствуют демонстрации [59].

Таким образом, использование агентов позволяет строить сложные системы, комбинирующие БЯМ и внешние инструменты, что позволяет превосходить одиночные БЯМ.

Ограничения. Несмотря на прогресс в развитии БЯМ и появления БРМ, они имеют значительные ограничения, особенно при решении задач, требующих рассуждений.

Важным недостатком БРМ является феномен чрезмерного обдумывания (англ. *overthinking*), который проявляется в генерации избыточных цепочек рассуждений, что приводит к увеличению вычислительных затрат и снижению эффективности моделей [60]. Наиболее явно данный эффект проявляется при решении простых задач [61]. Для преодоления эффекта чрезмерного обдумывания используется различные модификации процесса дообучения (обучение на данных с разной длиной, обучение с подкреплением с наградой за длину [62] или адаптивному переключению в режим без рассуждений [63]) и использования (назначение бюджета токенов на основе оценки сложности задачи и принудительное завершение процесса рассуждения [61]).

Сам процесс рассуждений, который генерируется БРМ – не обязательно коррелирует с ответом модели, что демонстрирует недостаточность мониторинга цепочки рассуждений для ее контроля и анализа [64].

Дообучение БЯМ с использованием обучения с подкреплением (RL) и проверяемыми вознаграждения (RLVR) улучшает способность моделей выбирать эффективные пути рассуждения уже присутствующие в базовых моделях, а не дает новую способность к рассуждению [65]. Проведенное сравнение метрики $\text{pass}@k$ (измеряет вероятность того, что хотя бы один из k ответов будет правильным) базовой модели (Qwen2.5 (7B/14B/32B), LLaMA-3.1-8B для математических задач и DeepSeek-R1-DistillQwen-14B для генерации кода) и их дообученных версий (GRPO) показало, что хотя при низких значениях выборки ($\text{pass}@1$) модели обученные RLVR превосходят свои базовые версии, но все меняется по мере увеличения числа выборок ответа. Базовые модели не только догоняют, но и превосходят модели, дообученные при помощи RLVR на контрольных тестах на математическое рассуждение (MATH500, AIME24, Minerva, Olympiad) и генерацию кода (LiveCodeBench, HumanEval+) [65].

Отмечается, что «момент озарения» или «Aha Moment» (проявился при обучении с подкреплением модели DeepSeek-R1-Zero [33]), на самом деле проявляется у базовой модели, что позволяет сделать заключение о том, что данные языковые артефакты отражают элементы обучающей выборки, а не проявление новых когнитивных процессов и способностей благодаря RL [66].

Сравнение контролируемого дообучения (SFT) и обучение с подкреплением показывает, что по сравнению с RL, модели дообученные рассуждению на математических задачах при помощи SFT не обобщают полученную способность на задачи из других областей [34]. Причем метод дообучения является более важным фактором, так как модели дообученные при помощи RL демонстрируют более высокий индекс переносимости, а дообучение моделей с использованием SFT может приводить к переобучению на математический домен, что приводит к снижению производительности в других областях. При этом, для понимания возможных причин различий в обобщающей способности было проведено сравнения главных компонент (англ. *Principal Component Analysis, PCA*), полученных из весовых значений, извлекаемых по слоям моделей. Сравнение величин сдвига (евклидово расстояние) PCA показало, что SFT сильнее приводит к изменению весов модели (что может приводить к ухудшению производительности), тогда как использование RL нарушает латентную структуру базовой модели в минимальной степени [34].

Исходя из успешного прохождения контрольных тестов, может сложиться впечатление, что БЯМ способны «понимать» концепции, которые скрываются за вопросами, но на самом деле – истинного осмысления у моделей не возникает [67]. Аналогичное заклю-

чение можно дать и рассуждающим моделям. Используя контрольные тесты, отличные от стандартных, было показано, что БРМ (OpenAI o1/o3, DeepSeek-R1, Claude 3.7 Sonnet Thinking, Gemini Thinking) менее эффективны (по сравнению со стандартными БЯМ) в решении простых задач, превосходят их на задачах средней сложности, но не справляются на сложных задачах (демонстрируя, при этом, снижение количества генерируемых токенов рассуждений) [68]. При этом, даже предоставление явных пошаговых алгоритмов решения используемых для проверки головоломок (Ханойская башня, переправа через реку и т.п.) не улучшает производительность моделей.

Несмотря на критику использованной методологии исследования, которая указывает на корреляцию обвала с достижением максимальной длины вывода, использования неразрешимых проблем (тест «переправа через реку» с $N \geq 6$, при вместимости лодки равной 3) и используемой методологии оценки правильности ответа [69], сама суть выводов подтверждается другим исследованием, которое так же демонстрирует «утомляемость» БРМ при решении сложных задач [70]. Проведенный анализ возникавших при решении задач затруднений, выявил: накопление ошибок с увеличением шагов рассуждений (пропуск информации, фактические ошибки), трудности с длинным контекстом (сложность выделения релевантных данных приводит к пропуску важной информации), статистические сокращения (в некоторых случаях модель может упрощать или делать догадки), недостаточное отслеживание состояния при выполнении многошаговых рассуждений (проявляется, например, в падении производительности при подсчете нескольких слов), плохая обобщающая способность вне распределения обучающих данных (OOD), ошибки копирования и проблемы с токенизацией (проявляется в задачах подсчета числа символов) [70]. Полученные результаты указывают, что БРМ в большей степени имитируют обучающие данные, а не выполняют «истинные рассуждения».

Таким образом, несмотря на успехи БРМ на стандартных тестах, их способность к рассуждению серьезно ограничена ростом сложности задач и глубины вывода. При этом, длинные цепочки рассуждений приводят к проблеме «чрезмерного обдумывания», лишним итерациям и некорректным выводам.

Заключение. В настоящее время большие языковые модели находятся на этапе интенсивного развития. Постоянно появляются новые модели, области и способы их использования. Однако базовая архитектура и процедура обучения БЯМ сводится к статистическому моделированию языка, что не наделяет БЯМ способностью рассуждения. При этом, именно способность к рассуждению является одним из важнейших аспектов систем ИИ.

В данной статье проведен обзор современных методов рассуждений, применяемых в больших языковых моделях в 2025 году, с акцентом на улучшение способности БЯМ к рассуждению и построению смысловых цепочек. Рассмотренные подходы, такие как построение цепочек мыслей, вычисления во время тестирования и интеграция внешних инструментов, позволяют моделям более эффективно справляться с многоступенчатыми задачами.

Для дальнейшего повышения интеллектуальных возможностей БЯМ требуется развитие данных подходов, включая внедрение модульной архитектуры, дообучение на синтетических данных, обучение с подкреплением и развитие многоагентского подхода. Данные направления являются актуальными для дальнейших исследований и представляют собой перспективные направления как для теоретического, так и для практического улучшения больших языковых моделей.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Vaswani A. Attention is all you need // *Advances in Neural Information Processing Systems*. – 2017.
2. Mirzadeh I., Alizadeh K., Shahrokhi H., Tuzel O., Bengio S., Farajtabar M. Gsm-symbolic: Understanding the limitations of mathematical reasoning in large language models // *arXiv preprint arXiv:2410.05229*. – 2024.
3. Minaee S., Mikolov T., Nikzad N., Chenaghlu M., Socher R., Amatriain X., Gao J. Large language models: A survey // *arXiv preprint arXiv:2402.06196*. – 2024.

4. Berglund L., Tong M., Kaufmann M., Balesni M., Stickland A. C., Korbak T., Evans O. The reversal curse: LLMs trained on "a is b" fail to learn "b is a" // arXiv preprint arXiv:2309.12288. – 2023.
5. Wu Z., Qiu L., Ross A., Akyürek E., Chen B., Wang B., Kim N., Andreas J., Kim Y. Reasoning or reciting? exploring the capabilities and limitations of language models through counterfactual tasks // arXiv preprint arXiv:2307.02477. – 2023.
6. Wang L., Ma C., Feng X., Zhang Z., Yang H., Zhang J., Chen Z., Tang J., Chen X., Lin Y., others. A survey on large language model based autonomous agents // *Frontiers of Computer Science*. – 2024. – Vol. 18, No. 6. – P. 186345.
7. Arkoudas K. GPT-4 can't reason // arXiv preprint arXiv:2308.03762. – 2023.
8. Chang Y., Wang X., Wang J., Wu Y., Yang L., Zhu K., Chen H., Yi X., Wang C., Wang Y., Ye W., Zhang Y., Chang Y., Yu P. S., Yang Q., Xie X. A Survey on Evaluation of Large Language Models // *ACM Trans. Intell. Syst. Technol.* – 2024. – Vol. 15, No. 3. – DOI: 10.1145/3641289.
9. Mahowald K., Ivanova A.A., Blank I. A., Kanwisher N., Tenenbaum J.B., Fedorenko E. Dissociating language and thought in large language models // *Trends in Cognitive Sciences*. – 2024.
10. Plaat A., Wong A., Verberne S., Broekens J., Stein N. van, Back T. Reasoning with large language models, a survey // arXiv preprint arXiv:2407.11511. – 2024.
11. Penedo G., Kydlíček H., Lozhkov A., Mitchell M., Raffel C., Von Werra L., Wolf T., others. The fineweb datasets: Decanting the web for the finest text data at scale // arXiv preprint arXiv:2406.17557. – 2024.
12. Shao Z., Wang P., Zhu Q., Xu R., Song J., Bi X., Zhang H., Zhang M., Li Y., Wu Y., others. Deepseekmath: Pushing the limits of mathematical reasoning in open language models // arXiv preprint arXiv:2402.03300. – 2024.
13. Aryabumi V., Su Y., Ma R., Morisot A., Zhang I., Locatelli A., Fadaee M., Üstün A., Hooker S. To Code, or Not To Code? Exploring Impact of Code in Pre-training // arXiv preprint arXiv:2408.10914. – 2024.
14. Morishita T., Morio G., Yamaguchi A., Sogawa Y. Enhancing Reasoning Capabilities of LLMs via Principled Synthetic Logic Corpus // arXiv preprint arXiv:2411.12498. – 2024.
15. Abdin M., Aneja J., Behl H., Bubeck S., Eldan R., Gunasekar S., Harrison M., Hewett R. J., Javaheripi M., Kauffmann P., others. Phi-4 Technical Report // arXiv preprint arXiv:2412.08905. – 2024.
16. Chen A., Li A., Gong B., Jiang B., Fei B., Yang B., Shan B., Yu C., Wang C., Zhu C., others. MiniMax-M1: Scaling Test-Time Compute Efficiently with Lightning Attention // arXiv preprint arXiv:2506.13585. – 2025.
17. Veličković P., Perivolaropoulos C., Barbero F., Pascanu R. Softmax is not enough (for sharp out-of-distribution) // arXiv preprint arXiv:2410.01104. – 2024.
18. Zhang Y., Backurs A., Bubeck S., Eldan R., Gunasekar S., Wagner T. Unveiling transformers with lego: a synthetic reasoning task // arXiv preprint arXiv:2206.04301. – 2022.
19. Li C., Liu J., Chen Y., Jia Y., Li Z. KunlunBaize: LLM with Multi-Scale Convolution and Multi-Token Prediction Under TransformerX Framework // arXiv preprint arXiv:2503.04784. – 2025.
20. Hao S., Sukhbaatar S., Su D., Li X., Hu Z., Weston J., Tian Y. Training large language models to reason in a continuous latent space // arXiv preprint arXiv:2412.06769. – 2024.
21. Зайцев Д.В. Почему большие языковые модели не (всегда) рассуждают как люди? // *Вестник Московского университета. Серия 7. Философия*. – 2024. – № 1. – С. 76-93.
22. Pagnoni A., Pasunuru R., Rodriguez P., Nguyen J., Muller B., Li M., Zhou C., Yu L., Weston J., Zettlemoyer L., others. Byte Latent Transformer: Patches Scale Better Than Tokens // arXiv preprint arXiv:2412.09871. – 2024.
23. Xia B., Shen B., Zhu D., Zhang D., Wang G., Zhang H., Liu H., Xiao J., Dong J., Zhao L., others. MiMo: Unlocking the Reasoning Potential of Language Model–From Pretraining to Posttraining // arXiv preprint arXiv:2505.07608. – 2025.
24. Zelikman E., Wu Y., Mu J., Goodman N. Star: Bootstrapping reasoning with reasoning // *Advances in Neural Information Processing Systems*. – 2022. – Vol. 35. – P. 15476-15488.
25. Wu T., Lan J., Yuan W., Jiao J., Weston J., Sukhbaatar S. Thinking LLMs: General Instruction Following with Thought Generation // arXiv preprint arXiv:2410.10630. – 2024.
26. Li W.-D., Hu K., Larsen C., Wu Y., Alford S., Woo C., Dunn S. M., Tang H., Naim M., Nguyen D., others. Combining induction and transduction for abstract reasoning // arXiv preprint arXiv:2411.02272. – 2024.
27. Kumar A., Zhuang V., Agarwal R., Su Y., Co-Reyes J. D., Singh A., Baumli K., Iqbal S., Bishop C., Roelofs R., others. Training language models to self-correct via reinforcement learning // arXiv preprint arXiv:2409.12917. – 2024.
28. Chen J. C.-Y., Wang Z., Palangi H., Han R., Ebrahimi S., Le L., Perot V., Mishra S., Bansal M., Lee C.-Y., others. Reverse Thinking Makes LLMs Stronger Reasoners // arXiv preprint arXiv:2411.19865. – 2024.

29. Muennighoff N., Yang Z., Shi W., Li X. L., Fei-Fei L., Hajishirzi H., Zettlemoyer L., Liang P., Candès E., Hashimoto T. s1: Simple test-time scaling // arXiv preprint arXiv:2501.19393. – 2025.
30. Ye Y., Huang Z., Xiao Y., Chern E., Xia S., Liu P. LIMO: Less is More for Reasoning. – 2025.
31. Learning to Reason with LLMs. – URL: <https://openai.com/index/learning-to-reason-with-llms/> (дата обращения: 22.11.2024).
32. Zeng Z., Cheng Q., Yin Z., Wang B., Li S., Zhou Y., Guo Q., Huang X., Qiu X. Scaling of Search and Learning: A Roadmap to Reproduce o1 from Reinforcement Learning Perspective. – 2024.
33. Guo D., Yang D., Zhang H., Song J., Zhang R., Xu R., Zhu Q., Ma S., Wang P., Bi X., others. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning // arXiv preprint arXiv:2501.12948. – 2025.
34. Huan M., Li Y., Zheng T., Xu X., Kim S., Du M., Poovendran R., Neubig G., Yue X. Does Math Reasoning Improve General LLM Capabilities? Understanding Transferability of LLM Reasoning. – 2025.
35. Wei J., Wang X., Schuurmans D., Bosma M., Xia F., Chi E., Le Q. V., Zhou D., others. Chain-of-thought prompting elicits reasoning in large language models // Advances in neural information processing systems. – 2022. – Vol. 35. – P. 24824-24837.
36. Kojima T., Gu S.S., Reid M., Matsuo Y., Iwasawa Y. Large language models are zero-shot reasoners // Advances in neural information processing systems. – 2022. – Vol. 35. – P. 22199-22213.
37. Long J. Large language model guided tree-of-thought // arXiv preprint arXiv:2305.08291. – 2023.
38. Besta M., Blach N., Kubicek A., Gerstenberger R., Podstawski M., Gianinazzi L., Gajda J., Lehmann T., Niewiadomski H., Nyczyk P., others. Graph of thoughts: Solving elaborate problems with large language models. – 2024. – P. 17682-17690.
39. Wang X., Wei J., Schuurmans D., Le Q., Chi E., Narang S., Chowdhery A., Zhou D. Self-consistency improves chain of thought reasoning in language models // arXiv preprint arXiv:2203.11171. – 2022.
40. Madaan A., Tandon N., Gupta P., Hallinan S., Gao L., Wiegrefe S., Alon U., Dziri N., Prabhume S., Yang Y., others. Self-refine: Iterative refinement with self-feedback // Advances in Neural Information Processing Systems. – 2024. – Vol. 36.
41. Miao N., Teh Y.W., Rainforth T. Selfcheck: Using llms to zero-shot check their own step-by-step reasoning // arXiv preprint arXiv:2308.00436. – 2023.
42. Shinn N., Cassano F., Gopinath A., Narasimhan K., Yao S. Reflexion: Language agents with verbal reinforcement learning // Advances in Neural Information Processing Systems. – 2024. – Vol. 36.
43. Lewis P., Perez E., Piktus A., Petroni F., Karpukhin V., Goyal N., Küttler H., Lewis M., Yih W., Rocktäschel T., others. Retrieval-augmented generation for knowledge-intensive nlp tasks // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – P. 9459-9474.
44. Jiang J., Chen J., Li J., Ren R., Wang S., Zhao W. X., Song Y., Zhang T. RAG-Star: Enhancing Deliberative Reasoning with Retrieval Augmented Verification and Refinement // arXiv preprint arXiv:2412.12881. – 2024.
45. Zhou P., Pujara J., Ren X., Chen X., Cheng H.-T., Le Q. V., Chi E. H., Zhou D., Mishra S., Zheng H. S. Self-discover: Large language models self-compose reasoning structures // arXiv preprint arXiv:2402.03620. – 2024.
46. Akyürek E., Damani M., Qiu L., Guo H., Kim Y., Andreas J. The Surprising Effectiveness of Test-Time Training for Abstract Reasoning // arXiv preprint arXiv:2411.07279. – 2024.
47. Snell C., Lee J., Xu K., Kumar A. Scaling llm test-time compute optimally can be more effective than scaling model parameters // arXiv preprint arXiv:2408.03314. – 2024.
48. Zhong T., Liu Z., Pan Y., Zhang Y., Zhou Y., Liang S., Wu Z., Lyu Y., Shu P., Yu X., others. Evaluation of openai o1: Opportunities and challenges of agi // arXiv preprint arXiv:2409.18486. – 2024.
49. Zhao Y., Yin H., Zeng B., Wang H., Shi T., Lyu C., Wang L., Luo W., Zhang K. Marco-o1: Towards Open Reasoning Models for Open-Ended Solutions // arXiv preprint arXiv:2411.14405. – 2024.
50. Scaling test-time compute - a Hugging Face Space by HuggingFaceH4. – URL: <https://huggingface.co/spaces/HuggingFaceH4/blogpost-scaling-test-time-compute> (дата обращения: 17.12.2024).
51. Zheng C., Zhang Z., Zhang B., Lin R., Lu K., Yu B., Liu D., Zhou J., Lin J. ProcessBench: Identifying Process Errors in Mathematical Reasoning // arXiv preprint arXiv:2412.06559. – 2024.
52. Schick T., Dwivedi-Yu J., Dessì R., Raileanu R., Lomeli M., Hambro E., Zettlemoyer L., Cancedda N., Scialom T. Toolformer: Language models can teach themselves to use tools // Advances in Neural Information Processing Systems. – 2023. – Vol. 36. – P. 68539-68551.
53. Li C., Liang J., Zeng A., Chen X., Hausman K., Sadigh D., Levine S., Fei-Fei L., Xia F., Ichter B. Chain of code: Reasoning with a language model-augmented code emulator // arXiv preprint arXiv:2312.04474. – 2023.
54. Chen W., Ma X., Wang X., Cohen W.W. Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks // arXiv preprint arXiv:2211.12588. – 2022.

55. Gao L., Madaan A., Zhou S., Alon U., Liu P., Yang Y., Callan J., Neubig G. Pal: Program-aided language models. – 2023. – P. 10764-10799.
56. Yao S., Zhao J., Yu D., Du N., Shafran I., Narasimhan K., Cao Y. React: Synergizing reasoning and acting in language models // arXiv preprint arXiv:2210.03629. – 2022.
57. Gou Z., Shao Z., Gong Y., Shen Y., Yang Y., Huang M., Duan N., Chen W. Tora: A tool-integrated reasoning agent for mathematical problem solving // arXiv preprint arXiv:2309.17452. – 2023.
58. Motwani S. R., Smith C., Das R. J., Rybchuk M., Torr P.H., Laptev I., Pizzati F., Clark R., Witt C.S. de MALT: Improving Reasoning with Multi-Agent LLM Training // arXiv preprint arXiv:2412.01928. – 2024.
59. Wang Q., Wang Z., Su Y., Tong H., Song Y. Rethinking the Bounds of LLM Reasoning: Are Multi-Agent Discussions the Key? // arXiv preprint arXiv:2402.18272. – 2024.
60. Chen X., Xu J., Liang T., He Z., Pang J., Yu D., Song L., Liu Q., Zhou M., Zhang Z., others. Do not think that much for $2+3=?$ on the overthinking of o1-like llms // arXiv preprint arXiv:2412.21187. – 2024.
61. Pu X., Saxon M., Hua W., Wang W.Y. THOUGHTTERMINATOR: Benchmarking, Calibrating, and Mitigating Overthinking in Reasoning Models // arXiv preprint arXiv:2504.13367. – 2025.
62. Sui Y., Chuang Y.-N., Wang G., Zhang J., Zhang T., Yuan J., Liu H., Wen A., Chen H., Hu X., others. Stop Overthinking: A Survey on Efficient Reasoning for Large Language Models // arXiv preprint arXiv:2503.16419. – 2025.
63. Fang G., Ma X., Wang X. Thinkless: LLM Learns When to Think // arXiv preprint arXiv:2505.13379. – 2025.
64. Chen Y., Benton J., Radhakrishnan A., Uesato J., Denison C., Schulman J., Somani A., Hase P., Wagner M., Roger F., others. Reasoning Models Don't Always Say What They Think // arXiv preprint arXiv:2505.05410. – 2025.
65. Yue Y., Chen Z., Lu R., Zhao A., Wang Z., Song S., Huang G. Does Reinforcement Learning Really Incentivize Reasoning Capacity in LLMs Beyond the Base Model? // arXiv preprint arXiv:2504.13837. – 2025.
66. Liu Z., Chen C., Li W., Pang T., Du C., Lin M. There May Not be Aha Moment in R1-Zero-like Training — A Pilot Study. – 2025.
67. Mancoridis M., Weeks B., Vafa K., Mullainathan S. Potemkin Understanding in Large Language Models // arXiv preprint arXiv:2506.21521. – 2025.
68. Shojaei P., Mirzadeh I., Alizadeh K., Horton M., Bengio S., Farajtabar M. The illusion of thinking: Understanding the strengths and limitations of reasoning models via the lens of problem complexity // arXiv preprint arXiv:2506.06941. – 2025.
69. Opus C., Lawsen A. The Illusion of the Illusion of Thinking // arXiv preprint ArXiv:2506.09250. – 2025.
70. Malek A., Ge J., Jin C., Györfy A., Szepesvári C. Frontier LLMs Still Struggle with Simple Reasoning Tasks // arXiv preprint arXiv:2507.07313. – 2025.

REFERENCES

1. Vaswani A. Attention is all you need, *Advances in Neural Information Processing Systems*, 2017.
2. Mirzadeh I., Alizadeh K., Shahrokhi H., Tuzel O., Bengio S., Farajtabar M. Gsm-symbolic: Understanding the limitations of mathematical reasoning in large language models, *arXiv preprint arXiv:2410.05229*, 2024.
3. Minaee S., Mikolov T., Nikzad N., Chenaghlu M., Socher R., Amatriain X., Gao J. Large language models: A survey, *arXiv preprint arXiv:2402.06196*, 2024.
4. Berglund L., Tong M., Kaufmann M., Balesni M., Stickland A. C., Korbak T., Evans O. The reversal curse: LLMs trained on "a is b" fail to learn "b is a", *arXiv preprint arXiv:2309.12288*, 2023.
5. Wu Z., Qiu L., Ross A., Akyürek E., Chen B., Wang B., Kim N., Andreas J., Kim Y. Reasoning or reciting? exploring the capabilities and limitations of language models through counterfactual tasks, *arXiv preprint arXiv:2307.02477*, 2023.
6. Wang L., Ma C., Feng X., Zhang Z., Yang H., Zhang J., Chen Z., Tang J., Chen X., Lin Y., others. A survey on large language model based autonomous agents, *Frontiers of Computer Science*, 2024, Vol. 18, No. 6, pp. 186345.
7. Arkoudas K. GPT-4 can't reason, *arXiv preprint arXiv:2308.03762*, 2023.
8. Chang Y., Wang X., Wang J., Wu Y., Yang L., Zhu K., Chen H., Yi X., Wang C., Wang Y., Ye W., Zhang Y., Chang Y., Yu P. S., Yang Q., Xie X. A Survey on Evaluation of Large Language Models, *ACM Trans. Intell. Syst. Technol.*, 2024, Vol. 15, No. 3. DOI: 10.1145/3641289.
9. Mahowald K., Ivanova A.A., Blank I.A., Kanwisher N., Tenenbaum J.B., Fedorenko E. Dissociating language and thought in large language models, *Trends in Cognitive Sciences*, 2024.

10. *Plaat A., Wong A., Verberne S., Broekens J., Stein N. van, Back T.* Reasoning with large language models, a survey, *arXiv preprint arXiv:2407.11511*, 2024.
11. *Penedo G., Kydliček H., Lozhkov A., Mitchell M., Raffel C., Von Werra L., Wolf T., others.* The fineweb datasets: Decanting the web for the finest text data at scale, *arXiv preprint arXiv:2406.17557*, 2024.
12. *Shao Z., Wang P., Zhu Q., Xu R., Song J., Bi X., Zhang H., Zhang M., Li Y., Wu Y., others.* Deepseekmath: Pushing the limits of mathematical reasoning in open language models, *arXiv preprint arXiv:2402.03300*, 2024.
13. *Aryabumi V., Su Y., Ma R., Morisot A., Zhang I., Locatelli A., Fadaee M., Üstün A., Hooker S.* To Code, or Not To Code? Exploring Impact of Code in Pre-training, *arXiv preprint arXiv:2408.10914*, 2024.
14. *Morishita T., Morio G., Yamaguchi A., Sogawa Y.* Enhancing Reasoning Capabilities of LLMs via Principled Synthetic Logic Corpus, *arXiv preprint arXiv:2411.12498*, 2024.
15. *Abdin M., Aneja J., Behl H., Bubeck S., Eldan R., Gunasekar S., Harrison M., Hewett R. J., Javaheripi M., Kauffmann P., others.* Phi-4 Technical Report, *arXiv preprint arXiv:2412.08905*, 2024.
16. *Chen A., Li A., Gong B., Jiang B., Fei B., Yang B., Shan B., Yu C., Wang C., Zhu C., others.* MiniMax-M1: Scaling Test-Time Compute Efficiently with Lightning Attention, *arXiv preprint arXiv:2506.13585*, 2025.
17. *Veličković P., Perivolaropoulos C., Barbero F., Pascanu R.* Softmax is not enough (for sharp out-of-distribution), *arXiv preprint arXiv:2410.01104*, 2024.
18. *Zhang Y., Backurs A., Bubeck S., Eldan R., Gunasekar S., Wagner T.* Unveiling transformers with lego: a synthetic reasoning task, *arXiv preprint arXiv:2206.04301*, 2022.
19. *Li C., Liu J., Chen Y., Jia Y., Li Z.* KunlunBaize: LLM with Multi-Scale Convolution and Multi-Token Prediction Under TransformerX Framework, *arXiv preprint arXiv:2503.04784*, 2025.
20. *Hao S., Sukhbaatar S., Su D., Li X., Hu Z., Weston J., Tian Y.* Training large language models to reason in a continuous latent space, *arXiv preprint arXiv:2412.06769*, 2024.
21. *Zaytsev D.V.* Pochemu bol'shie yazykovye modeli ne (vsegda) rassuzhdayut kak lyudi? [Why don't large language models (always) reason like humans?], *Vestnik Moskovskogo universiteta. Seriya 7. Filosofiya* [Moscow University Bulletin. Series 7. Philosophy], 2024, No. 1, pp. 76-93.
22. *Pagnoni A., Pasunuru R., Rodriguez P., Nguyen J., Muller B., Li M., Zhou C., Yu L., Weston J., Zettlemoyer L., others.* Byte Latent Transformer: Patches Scale Better Than Tokens, *arXiv preprint arXiv:2412.09871*, 2024.
23. *Xia B., Shen B., Zhu D., Zhang D., Wang G., Zhang H., Liu H., Xiao J., Dong J., Zhao L., others.* MiMo: Unlocking the Reasoning Potential of Language Model-From Pretraining to Posttraining, *arXiv preprint arXiv:2505.07608*, 2025.
24. *Zelikman E., Wu Y., Mu J., Goodman N.* Star: Bootstrapping reasoning with reasoning, *Advances in Neural Information Processing Systems*, 2022, Vol. 35, pp. 15476-15488.
25. *Wu T., Lan J., Yuan W., Jiao J., Weston J., Sukhbaatar S.* Thinking LLMs: General Instruction Following with Thought Generation // *arXiv preprint arXiv:2410.10630*. – 2024.
26. *Li W.-D., Hu K., Larsen C., Wu Y., Alford S., Woo C., Dunn S. M., Tang H., Naim M., Nguyen D., others.* Combining induction and transduction for abstract reasoning, *arXiv preprint arXiv:2411.02272*, 2024.
27. *Kumar A., Zhuang V., Agarwal R., Su Y., Co-Reyes J. D., Singh A., Baumli K., Iqbal S., Bishop C., Roelofs R., others.* Training language models to self-correct via reinforcement learning, *arXiv preprint arXiv:2409.12917*, 2024.
28. *Chen J. C.-Y., Wang Z., Palangi H., Han R., Ebrahimi S., Le L., Perot V., Mishra S., Bansal M., Lee C.-Y., others.* Reverse Thinking Makes LLMs Stronger Reasoners, *arXiv preprint arXiv:2411.19865*, 2024.
29. *Muennighoff N., Yang Z., Shi W., Li X. L., Fei-Fei L., Hajishirzi H., Zettlemoyer L., Liang P., Candès E., Hashimoto T. s1:* Simple test-time scaling, *arXiv preprint arXiv:2501.19393*, 2025.
30. *Ye Y., Huang Z., Xiao Y., Chern E., Xia S., Liu P.* LIMO: Less is More for Reasoning, 2025.
31. *Learning to Reason with LLMs.* Available at: <https://openai.com/index/learning-to-reason-with-llms/> (accessed 22 November 2024).
32. *Zeng Z., Cheng Q., Yin Z., Wang B., Li S., Zhou Y., Guo Q., Huang X., Qiu X.* Scaling of Search and Learning: A Roadmap to Reproduce o1 from Reinforcement Learning Perspective, 2024.
33. *Guo D., Yang D., Zhang H., Song J., Zhang R., Xu R., Zhu Q., Ma S., Wang P., Bi X., others.* DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning, *arXiv preprint arXiv:2501.12948*, 2025.
34. *Huan M., Li Y., Zheng T., Xu X., Kim S., Du M., Poovendran R., Neubig G., Yue X.* Does Math Reasoning Improve General LLM Capabilities? Understanding Transferability of LLM Reasoning, 2025.

35. Wei J., Wang X., Schuurmans D., Bosma M., Xia F., Chi E., Le Q. V., Zhou D., others. Chain-of-thought prompting elicits reasoning in large language models, *Advances in neural information processing systems*, 2022, Vol. 35, pp. 24824-24837.
36. Kojima T., Gu S.S., Reid M., Matsuo Y., Iwasawa Y. Large language models are zero-shot reasoners, *Advances in neural information processing systems*, 2022, Vol. 35, pp. 22199-22213.
37. Long J. Large language model guided tree-of-thought, *arXiv preprint arXiv:2305.08291*, 2023.
38. Besta M., Blach N., Kubicek A., Gerstenberger R., Podstawski M., Gianinazzi L., Gajda J., Lehmann T., Niewiadomski H., Nyczyk P., others. Graph of thoughts: Solving elaborate problems with large language models, 2024, pp. 17682-17690.
39. Wang X., Wei J., Schuurmans D., Le Q., Chi E., Narang S., Chowdhery A., Zhou D. Self-consistency improves chain of thought reasoning in language models, *arXiv preprint arXiv:2203.11171*, 2022.
40. Madaan A., Tandon N., Gupta P., Hallinan S., Gao L., Wiegrefe S., Alon U., Dziri N., Prabhume S., Yang Y., others. Self-refine: Iterative refinement with self-feedback, *Advances in Neural Information Processing Systems*, 2024, Vol. 36.
41. Miao N., Teh Y.W., Rainforth T. Selfcheck: Using llms to zero-shot check their own step-by-step reasoning, *arXiv preprint arXiv:2308.00436*, 2023.
42. Shinn N., Cassano F., Gopinath A., Narasimhan K., Yao S. Reflexion: Language agents with verbal reinforcement learning, *Advances in Neural Information Processing Systems*, 2024, Vol. 36.
43. Lewis P., Perez E., Piktus A., Petroni F., Karpukhin V., Goyal N., Küttler H., Lewis M., Yih W., Rocktäschel T., others. Retrieval-augmented generation for knowledge-intensive nlp tasks, *Advances in Neural Information Processing Systems*, 2020, Vol. 33, pp. 9459-9474.
44. Jiang J., Chen J., Li J., Ren R., Wang S., Zhao W. X., Song Y., Zhang T. RAG-Star: Enhancing Deliberative Reasoning with Retrieval Augmented Verification and Refinement, *arXiv preprint arXiv:2412.12881*, 2024.
45. Zhou P., Pujara J., Ren X., Chen X., Cheng H.-T., Le Q.V., Chi E.H., Zhou D., Mishra S., Zheng H.S. Self-discover: Large language models self-compose reasoning structures, *arXiv preprint arXiv:2402.03620*, 2024.
46. Akyürek E., Damani M., Qiu L., Guo H., Kim Y., Andreas J. The Surprising Effectiveness of Test-Time Training for Abstract Reasoning, *arXiv preprint arXiv:2411.07279*, 2024.
47. Snell C., Lee J., Xu K., Kumar A. Scaling llm test-time compute optimally can be more effective than scaling model parameters, *arXiv preprint arXiv:2408.03314*, 2024.
48. Zhong T., Liu Z., Pan Y., Zhang Y., Zhou Y., Liang S., Wu Z., Lyu Y., Shu P., Yu X., others. Evaluation of openai o1: Opportunities and challenges of agi, *arXiv preprint arXiv:2409.18486*, 2024.
49. Zhao Y., Yin H., Zeng B., Wang H., Shi T., Lyu C., Wang L., Luo W., Zhang K. Marco-o1: Towards Open Reasoning Models for Open-Ended Solutions, *arXiv preprint arXiv:2411.14405*, 2024.
50. Scaling test-time compute - a Hugging Face Space by HuggingFaceH4. Available at: <https://huggingface.co/spaces/HuggingFaceH4/blogpost-scaling-test-time-compute> (accessed 17 Desember 2024).
51. Zheng C., Zhang Z., Zhang B., Lin R., Lu K., Yu B., Liu D., Zhou J., Lin J. ProcessBench: Identifying Process Errors in Mathematical Reasoning, *arXiv preprint arXiv:2412.06559*, 2024.
52. Schick T., Dwivedi-Yu J., Dessì R., Raileanu R., Lomeli M., Hambro E., Zettlemoyer L., Cancedda N., Scialom T. Toolformer: Language models can teach themselves to use tools, *Advances in Neural Information Processing Systems*, 2023, Vol. 36, pp. 68539-68551.
53. Li C., Liang J., Zeng A., Chen X., Hausman K., Sadigh D., Levine S., Fei-Fei L., Xia F., Ichter B. Chain of code: Reasoning with a language model-augmented code emulator, *arXiv preprint arXiv:2312.04474*, 2023.
54. Chen W., Ma X., Wang X., Cohen W.W. Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks, *arXiv preprint arXiv:2211.12588*, 2022.
55. Gao L., Madaan A., Zhou S., Alon U., Liu P., Yang Y., Callan J., Neubig G. Pal: Program-aided language models, 2023, pp. 10764-10799.
56. Yao S., Zhao J., Yu D., Du N., Shafran I., Narasimhan K., Cao Y. React: Synergizing reasoning and acting in language models, *arXiv preprint arXiv:2210.03629*, 2022.
57. Gou Z., Shao Z., Gong Y., Shen Y., Yang Y., Huang M., Duan N., Chen W. Tora: A tool-integrated reasoning agent for mathematical problem solving, *arXiv preprint arXiv:2309.17452*, 2023.
58. Motwani S. R., Smith C., Das R. J., Rybchuk M., Torr P.H., Laptev I., Pizzati F., Clark R., Witt C.S. de MALT: Improving Reasoning with Multi-Agent LLM Training, *arXiv preprint arXiv:2412.01928*, 2024.
59. Wang Q., Wang Z., Su Y., Tong H., Song Y. Rethinking the Bounds of LLM Reasoning: Are Multi-Agent Discussions the Key?, *arXiv preprint arXiv:2402.18272*, 2024.

60. Chen X., Xu J., Liang T., He Z., Pang J., Yu D., Song L., Liu Q., Zhou M., Zhang Z., others. Do not think that much for $2+3=?$ on the overthinking of o1-like llms, *arXiv preprint arXiv:2412.21187*, 2024.
61. Pu X., Saxon M., Hua W., Wang W.Y. THOUGHTTERMINATOR: Benchmarking, Calibrating, and Mitigating Overthinking in Reasoning Models, *arXiv preprint arXiv:2504.13367*, 2025.
62. Sui Y., Chuang Y.-N., Wang G., Zhang J., Zhang T., Yuan J., Liu H., Wen A., Chen H., Hu X., others. Stop Overthinking: A Survey on Efficient Reasoning for Large Language Models, *arXiv preprint arXiv:2503.16419*, 2025.
63. Fang G., Ma X., Wang X. Thinkless: LLM Learns When to Think, *arXiv preprint arXiv:2505.13379*, 2025.
64. Chen Y., Benton J., Radhakrishnan A., Uesato J., Denison C., Schulman J., Somani A., Hase P., Wagner M., Roger F., others. Reasoning Models Don't Always Say What They Think, *arXiv preprint arXiv:2505.05410*, 2025.
65. Yue Y., Chen Z., Lu R., Zhao A., Wang Z., Song S., Huang G. Does Reinforcement Learning Really Incentivize Reasoning Capacity in LLMs Beyond the Base Model?, *arXiv preprint arXiv:2504.13837*, 2025.
66. Liu Z., Chen C., Li W., Pang T., Du C., Lin M. There May Not be Aha Moment in R1-Zero-like Training — A Pilot Study, 2025.
67. Mancoridis M., Weeks B., Vafa K., Mullainathan S. Potemkin Understanding in Large Language Models, *arXiv preprint arXiv:2506.21521*, 2025.
68. Shojae P., Mirzadeh I., Alizadeh K., Horton M., Bengio S., Farajtabar M. The illusion of thinking: Understanding the strengths and limitations of reasoning models via the lens of problem complexity, *arXiv preprint arXiv:2506.06941*, 2025.
69. Opus C., Lawsen A. The Illusion of the Illusion of Thinking, *arXiv preprint arXiv:2506.09250*, 2025.
70. Malek A., Ge J., Jin C., György A., Szepesvári C. Frontier LLMs Still Struggle with Simple Reasoning Tasks, *arXiv preprint arXiv:2507.07313*, 2025.

Савинов Владимир Борисович – Балтийский федеральный университет имени Иммануила Канта; e-mail: vsavinov@kantiana.ru; г. Калининград, Россия; Балтийский центр нейротехнологий и искусственного интеллекта; аспирант.

Шушарина Наталья Николаевна – Балтийский федеральный университет имени Иммануила Канта; e-mail: nshusharina@kantiana.ru; г. Калининград, Россия; Балтийский центр нейротехнологий и искусственного интеллекта; к. пед. н.; начальник управления развития и инновационной деятельности; старший научный сотрудник.

Savinov Vladimir Borisovich – Immanuel Kant Baltic Federal University; e-mail: vsavinov@kantiana.ru; Kaliningrad, Russia; Center for Neurotechnologies and Artificial Intelligence; postgraduate student.

Shusharina Natalia Nikolaevna – Immanuel Kant Baltic Federal University; e-mail: nshusharina@kantiana.ru; Kaliningrad, Russia; Baltic Center for Neurotechnologies and Artificial Intelligence; cand. of ped. sc.; head of the Department of Development and Innovation Activity; senior researcher.

УДК 004.722

DOI 10.18522/2311-3103-2026-1-270-284

А.В. Бобряков, С.А. Прокопенко

РАННЕЕ ВЫЯВЛЕНИЕ ПРОИЗВОДСТВЕННОГО БРАКА НА МЕЛКОСЕРИЙНЫХ ПРОИЗВОДСТВАХ С ИСПОЛЬЗОВАНИЕМ НЕЙРО-НЕЧЕТКИХ СИСТЕМ

Постановка задачи: Повышение требований к качеству продукции на мелкосерийных производствах и сложность раннего выявления производственного брака актуализируют необходимость разработки инновационных подходов к прогнозированию и контролю дефектов на ранних этапах производства сложных технических объектов. Известные методы, применяемые в массовом производстве, не подходят для мелкосерийных производств вследствие высокой изменчивости технологических процессов и недостаточности данных для традиционного статистического анализа. **Целью работы** является снижение уровня производственного брака путем раннего выявления отклонений на подготовительных этапах производства. Предлагается использование ней-

ро-нечетких систем, которые позволяют адаптивно прогнозировать дефекты на основе исторических данных. **Используемые методы:** для решения задачи раннего выявления производственного брака использованы нейро-нечеткие компоненты на основе нечеткого нейрона Квана–Кэи, позволяющие интегрировать экспертные знания с данными производства. Система включает подсистему обучения и тонкой настройки, состоящую из модулей подготовки, валидации и нормализации данных, фазификации данных и вычисления ошибки прогноза. В качестве структурного элемента прогнозирования применены темпоральные нейро-нечеткие сети Петри, которые обеспечивают учет временных аспектов и неопределенности в производственных процессах. **Новизна:** к элементам новизны относятся использование темпоральных нейро-нечетких сетей Петри и нейро-нечетких компонентов на основе нечеткого нейрона Квана–Кэи, позволяющих производить раннюю диагностику брака и принимать упреждающие меры. Новым является также подход интеграции нейро-нечетких методов с существующими системами управления производством. **Результат:** внедрение предложенных методов позволило снизить процент производственного брака на 15% за счет раннего выявления отклонений и своевременного принятия корректирующих мер. Реализованные программно-инструментальные средства обеспечивают оперативный анализ производственной ситуации и прогнозирование дефектов в режиме, близком к реальному времени. **Практическая значимость:** предлагаемое решение реализовано в виде специализированного программного обеспечения, интегрируемого в существующие производственные системы. Оно обеспечивает повышение эффективности управления качеством, снижает затраты на исправление дефектов и может быть адаптировано к различным мелкосерийным производствам, существенно улучшая показатели их работы.

Нейро-нечеткие компоненты; сети Петри; производственный брак.

A.V. Bobryakov, S.A. Prokopenko

EARLY DETECTION OF MANUFACTURING DEFECTS IN SMALL-SCALE PRODUCTION USING NEURO-FUZZY SYSTEMS

Problem Statement: The increasing demand for higher quality products in small-scale production and the complexity associated with the early detection of manufacturing defects necessitate the development of innovative approaches to predict and control defects at the early stages of manufacturing complex technical objects. Traditional methods applied in mass production settings are unsuitable for small-scale manufacturing due to the high variability of technological processes and the insufficient data required for conventional statistical analyses. **The objective of this study** is to reduce the incidence of manufacturing defects by identifying deviations at the preparatory stages of production. The proposed solution involves employing neuro-fuzzy systems capable of adaptively forecasting defects based on historical production data. **Methods:** To address the early detection of manufacturing defects, neuro-fuzzy components based on the fuzzy neuron proposed by Kwan–Cai were employed, integrating expert knowledge with production data. The system includes a training and fine-tuning subsystem consisting of modules for data preparation, validation and normalization, fuzzification of data, and calculation of forecasting errors. Temporal neuro-fuzzy Petri nets were used as structural forecasting elements, enabling the consideration of temporal aspects and uncertainties inherent in manufacturing processes. **Novelty:** The novel aspects of this research include the utilization of temporal neuro-fuzzy Petri nets and neuro-fuzzy components based on the Kwan–Cai fuzzy neuron, enabling the early detection of defects and the implementation of proactive measures. Another innovative aspect is the approach for integrating neuro-fuzzy methods into existing production management systems. **Results:** The implementation of the proposed methods resulted in a 15% reduction in manufacturing defects through early identification of deviations and the timely adoption of corrective actions. Developed software tools provide operational analysis of production situations and defect forecasting in near real-time. **Practical Significance:** The presented solution has been realized as specialized software integrated into existing production systems. It improves the effectiveness of quality management, reduces the costs associated with defect correction, and can be adapted to various small-scale production environments, significantly enhancing their operational performance.

Neuro-fuzzy components; Petri nets; manufacturing defects.

Введение. Целью данного исследования является изучение и разработка методов на основе нейро-нечетких систем для раннего выявления производственного брака и повышения эффективности управления качеством на мелкосерийных производствах. Актуальность темы обусловлена ростом требований к качеству продукции, сложностью процессов контроля и высокой стоимостью производственного брака. Особенно остро эта

проблема стоит на предприятиях, выпускающих продукцию малыми партиями с большим разнообразием номенклатуры. Даже незначительные дефекты могут приводить к серьёзным потерям, поэтому важно обнаруживать отклонения как можно раньше. Согласно правилу «1-10-100» [1], цена исправления дефекта многократно возрастает на каждой следующей стадии производства. Следовательно, эффективное раннее выявление брака способно существенно снизить издержки за счёт предотвращения переработки, простоя оборудования, переделок и рекламаций.

Массовые и крупносерийные производства традиционно используют множество инструментов управления качеством (статистический контроль процессов, шесть сигм и др.). Однако для мелкосерийного и единичного производства таких разработок значительно меньше. Основное отличие мелкосерийного производства – большая номенклатура изделий и частая смена уникальных технологических процессов, что затрудняет применение стандартных методов обеспечения качества. В условиях высокой изменчивости процессов трудно накопить достаточно данных для статистического контроля стабильности процессов, и требования к гибкости системы качества возрастают. Таким образом, возникает необходимость в инновационных подходах к мониторингу и прогнозированию качества, способных адаптироваться к новым условиям и небольшим объёмам однотипных данных.

В данной работе предложен подход, основанный на нейро-нечетких системах раннего обнаружения дефектов в производственном процессе. Нейро-нечеткие системы интегрируют способности искусственных нейронных сетей и нечеткой логики, сочетая обучаемость с человеческим стилем рассуждения на основе правил. Предлагаемые нейро-нечеткие компоненты прогнозирования брака обучаются на исторических данных и позволяют оценивать качество выполняемых операций до их завершения [2].

Производственный брак. Производственный брак – это продукция, изготовленная с отклонениями от установленных стандартов качества, не соответствующая техническим требованиям и не пригодная для использования по назначению. В процессе производства сложных технических объектов (СТО), включающем множество технологических операций, брак может возникать на разных этапах. На предприятиях применяются традиционные методы выявления брака, включая визуальный контроль и контроль качества на заключительных стадиях [3]. Однако такие подходы зачастую неэффективны для предупреждения дефектов на ранних этапах, особенно в условиях мелкосерийного и уникального производства, где каждый экземпляр изделия может проходить свою уникальную технологическую цепочку.

Основными причинами возникновения брака являются:

- ♦ дефекты в сырье, материалах, полуфабрикатах, поставленные внешними поставщиками;
- ♦ неисправность оборудования, вызванная неправильной наладкой оборудования и компонентов;
- ♦ ошибки в нормативно-технической документации;
- ♦ недостаточная квалификации рабочих;
- ♦ нарушение технологической дисциплины;
- ♦ чрезвычайное обстоятельство (стихийное бедствие).

По месту обнаружения различают внутренний брак (выявленный на предприятии при внутренних контрольных мероприятиях) и внешний брак (выявленный конечным потребителем при эксплуатации изделия). Внешний брак наиболее опасен и дорог для предприятия, так как обнаруживается поздно – когда продукт уже произведён и поставлен. Стоимость такого дефекта включает не только затраты на изготовление бракованного изделия, но и возможные штрафы, возвраты и снижение репутации предприятия.

Ключевые факторы, влияющие на соответствие СТО стандартам качества, можно разделить на несколько классов:

- ♦ оборудование – техническое состояние средств производства, правильность их настройки, износ и т.д.;

- ♦ персонал – профессиональное мастерство, знания, навыки и психофизиологическое состояние работников, соблюдение ими инструкций;
- ♦ сырьё и материалы – качество исходных материалов, полуфабрикатов, комплектующих, соответствие их спецификациям;
- ♦ методы и процесс – корректность технологической документации, адекватность выбранных режимов обработки, условия внешней среды.

Для массовых и серийных процессов разработаны многочисленные системы управления качеством, однако для мелкосерийных производств сложность состоит в постоянной смене продуктов и процессов. Каждая новая партия изделия может требовать новых настроек и контроля новых параметров, что затрудняет использование стандартных статистических методов. Необходим инструмент, способный быстро обучаться на небольших выборках данных, учитывать экспертные знания о процессе и работать в условиях неполной определённости.

Необходимо стремиться к выявлению потенциальных дефектов ещё до выполнения технологической операции – превентивно [4, 5]. Однако сделать это сложно, так как нужно предсказывать результат операции на основе совокупности факторов до её завершения. Для решения этой задачи могут быть использованы методы искусственного интеллекта, в частности нейро-нечеткие системы, способные обрабатывать большой набор входных параметров, неопределённости и взаимодействия факторов в реальном времени.

Нейро-нечеткие компоненты. Нейро-нечеткие системы – это гибрид искусственной нейронной сети и нечеткого логического вывода [6]. Такие системы сочетают обучаемость и способность аппроксимировать сложные зависимости, характерные для нейросетей, с прозрачностью и интерпретируемостью, присущей нечеткой логике (правила вида «если – то»). В контексте управления качеством нейро-нечеткие компоненты могут служить для прогнозирования появления дефектов на основе множества параметров процесса.

На ряде современных предприятий (автомобилестроение, аэрокосмическая отрасль, электроника и др.) уже успешно применяются нейро-нечеткие системы в задачах контроля качества. Внедрение таких систем позволяет своевременно выявлять потенциальные дефекты, оптимизировать технологические режимы и снижать затраты на бракованную продукцию [7]. Например, в литейном производстве колесных дисков предложен нейро-нечеткий классификатор дефектов рентгеновских снимков, показавший практически идеальную точность (площадь под ROC-кривой $A_z = 0.9976$) при разделении дефектов и ложных срабатываний [8, 9], что демонстрирует возможности нейро-нечетких подходов в задачах обнаружения брака.

Использование нейро-нечеткого компонента для раннего определения брака позволяет предсказать возможные дефекты ещё до выполнения технологической операции. То есть система, получив текущие параметры оборудования, материала и состояния процесса, может выдать предупреждение о вероятном браке, и тогда можно принять меры заблаговременно (подстроить режим, заменить инструмент, провести дополнительную настройку). Такой упреждающий контроль сокращает издержки на переработку бракованных деталей и повышает эффективность всего производственного цикла.

Нейро-нечеткие компоненты обладают рядом существенных преимуществ в задачах мониторинга качества.

1. Адаптивность и гибкость, достигающиеся за счёт обучения на исторических данных. Такая нейро-нечеткая система способна подстраиваться под меняющиеся условия производства и доступны возможности тонкой настройки модели при появлении новых видов продукции или изменении процесса, что делает её универсальным инструментом анализа качества.

2. Нейро-нечеткие системы являются универсальными аппроксиматорами [10] и могут моделировать практически любые нелинейные зависимости, сохраняя при этом интерпретируемость в виде нечетких правил.

3. Высокая точность и скорость обработки данных. Обученная нейро-нечеткая модель способна обрабатывать большие объёмы данных, характеризующих процессы, и выдавать решение в реальном или квазиреальном времени. Такая модель обнаруживает

скрытые многомерные зависимости между параметрами, обеспечивая более точное выявление потенциальных дефектов. В реальных экспериментах такие системы успешно классифицируют дефекты даже при большом количестве признаков, используя лишь наиболее информативные из них.

4. Нейро-нечеткий компонент обладает способностью к обучению на исторических данных. По мере пополнения базы данных система будет проводить тонкую самонастройку, улучшая точность прогнозов, автоматически подстраивая внутренние правила и параметры, что особенно важно в условиях мелкосерийного производства, где статистика ограничена – система способна извлечь максимум полезной информации даже из небольшого числа наблюдений.

5. Автоматизация процесса выявления брака с помощью нейро-нечетких алгоритмов минимизирует влияние человеческого фактора [11, 12], снижая влияние субъективности оператора. Система объективно оценивает параметры и выдает рекомендации, исключая «человеческий фактор» – усталость, невнимательность или разночтения технических требований. Роль человека смещается в сторону принятия решений по результатам работы интеллектуальной подсистемы, что повышает общую надежность контроля.

6. Экономический эффект представляет собой улучшение качества и снижение брака. Инвестиции в интеллектуальную систему окупаются за счёт сокращения потерь материала, времени и трудозатрат на переделки [13]. Повышение выхода годной продукции с первого раза увеличивает прибыль предприятия, кроме того, улучшается репутация за счёт стабильного качества, что косвенно влияет на экономику.

7. Нейро-нечеткие модули прогнозирования могут быть относительно легко интегрированы в существующие системы управления производством и контроля качества. Их можно встроить в АСУ ТП, SCADA или MES-системы предприятия, используя потоки данных, которые уже собираются. Благодаря модульности, подобные компоненты находят применение в различных отраслях без кардинальной перестройки действующей инфраструктуры.

Применение нейро-нечеткого компонента для предсказания брака требует не только наличия алгоритма, но и организации процесса обучения модели. Только при корректном обучении на качественных данных система будет надёжно работать.

Нечеткий нейрон Квана и Кэи. Нечеткий нейрон Квана и Кэи впервые был представлен в 1994 году. Он успешно применяется в задачах классификации [14]. Структурно он состоит из четырех слоев нейронов, предназначенных для фаззификации входных данных, анализа и классификации информации. Простота реализации и высокая точность классификации делают его эффективным инструментом в задачах прогнозирования брака. Схема нечеткого нейрона представлена на рис. 1.

Первый слой представляет входные данные и состоит из нечетких входов, каждый из которых представляет вектор параметров характеристик. В случае двумерной выборки, содержащей векторные элементы $N_1 \times N_2$, у нас будет первый слой с нейронами $N_1 \times N_2$. Для нейрона в позиции (i, j) уравнения имеют вид:

$$s_{ij}^{[1]} = z_{ij}^{[1]} = x_{ij},$$

$$y_{ij}^{[1]} = y_{ij}^{[1]} / P_{vmax},$$

где $i=1, 2, \dots, N_1$; $j=1, 2, \dots, N_2$, где $s_{ij}^{[1]}$ представляет состояние нейрона на (i, j) позиции для первого слоя; $z_{ij}^{[1]}$ – его входное значение; x_{ij} – значение элемента (i, j) во входном векторе выборки ($x_{ij} \geq 0$); $y_{ij}^{[1]}$ – его выходное значение; P_{vmax} – максимальное значение из всех входных элементов.

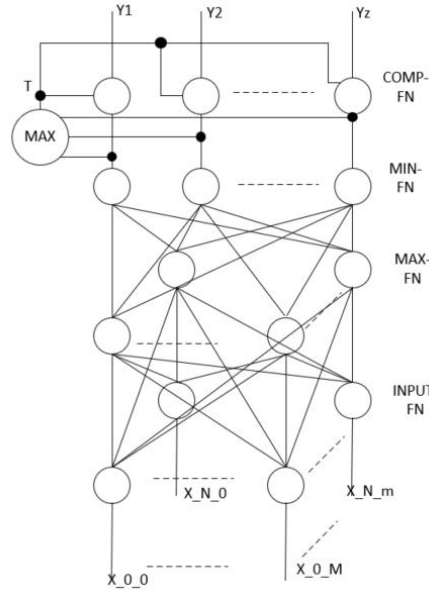


Рис. 1. Структура нечеткого нейрона Квана и Кэи

Второй слой построен из $N_1 \times N_2$ нейронов, и его задача – выполнить фаззификацию входных векторов с помощью весовой функции $w(m,n)$ – называемой функцией фаззификации, графически представленной на рис. 2 и определяемой как:

$$w(m,n) = e^{-\beta^2(m^2+n^2)},$$

где параметры $m = -(N_1-1), \dots, (N_1-1)$; $n = -(N_2-1), \dots, (N_2-1)$ и β определяют, какую большую часть выборочного вектора видит каждый нечеткий нейрон. Каждый нейрон второго слоя имеет M выходов, по одному на каждый нейрон третьего слоя. Выходные данные для нейрона второго слоя в позиции (p, q) :

$$y_{pqm}^{[2]} = q_{pqm}^{[2]},$$

где $p = 1, \dots, N_1$; $q = 1, \dots, N_2$; $m = 1, \dots, M$; $y_{pqm}^{[2]}$ m -выход нейрона второго слоя в позиции (p,q) к m -нейрону третьего уровня. Выходная функция q_{pqm} определяется в рамках алгоритма обучения. Для более упрощенного подхода использованы равнобедренные треугольники с основанием α и высотой 1, математически определяемые как:

$$y_{pqm}^{[2]} = q_{pqm}^{[2]}(s_{pq}^{[2]}) = \begin{cases} 1 - \frac{2|s_{pq}^{[2]} - \theta_{pqm}|}{\alpha}, & \text{for } 2|s_{pq}^{[2]} - \theta_{pqm}| \leq \frac{\alpha}{2}, \\ 0 & \text{other} \end{cases}$$

где $\alpha \geq 0$, $p = 1, \dots, N_1$; $q = 1, \dots, N_2$; $m = 1, \dots, M$. Параметр θ_{pqm} – центр основания равнобедренного треугольника. С помощью алгоритма обучения определяются значения p , q и m , соответствующие α и θ_{pqm} .

Третий слой состоит из M нейронов, где M – это количество определяемых классов. Его можно рассматривать как нечеткий слой вывода (вывода). Выходные данные нейрона третьего слоя:

$$y_m^{[3]} = s_m^{[3]} = \min_{p=1 \dots N_2} (y_{pqm}^{[2]}),$$

где $m = 1, \dots, M$.

Четвертый и последний слой является выходным слоем сети и состоит из конкурирующих нейронов, по одному на каждый из выходных классов – это слой дефаззификации. Если входной шаблон более похож на m -класс, который был изучен, то выходу m сравнительного нейрона будет присвоено значение 1, а остальным – значение 0:

$$y_m^{[4]} = s_m^{[4]} = z_m^{[4]},$$

$$y_{pqm}^{[4]} = q[s_m^{[4]} - T] = \begin{cases} 0, & \text{if } s_m^{[4]} < T \\ 1, & \text{if } s_m^{[4]} = T \end{cases},$$

$$T = \max_{m=1...M} (\max_{j=1...N_2} (y_m^{[3]}))$$

где $m=1,...,M$, где T определяется как порог активации для всех нейронов четвертого слоя.

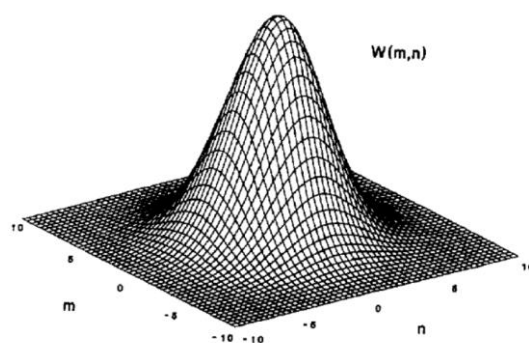


Рис. 2. Представление функции нечеткого нейрона Квана и Кэи

Нечеткий нейрон Квана–Кэи обеспечивает сквозной нечеткий вывод: от фаззификации входов через нечеткие правила к четкому решению о классе. Его важное свойство – возможность выдавать вектор выходов ($y_1...y_M$) с интерпретацией как значения функций принадлежности образца ко всем M классам. Это означает, что система не просто даёт «вердикт», но и предоставляет степень уверенности в каждом возможном исходе, что ценно для принятия решений (например, при промежуточных значениях можно направить операцию на дополнительный контроль).

В контексте задачи прогнозирования брака мы используем данную архитектуру следующим образом: на вход нейрона подаются параметры текущей производственной ситуации (характеристики оборудования, материалы, параметры процесса, данные об операторе и т.д.), а на выходе получаем принадлежность к классам «операция пройдет успешно», «возможны отклонения – нужен контроль» или «высокая вероятность брака». Далее эти результаты интерпретируются системой управления качеством для принятия мер.

Успешное применение таких компонентов требует их тщательной настройки и обучения на примерах.

Подсистема обучения нейро-нечеткого компонента. Для обеспечения работоспособности нейро-нечеткого компонента в реальных условиях разработана специальная подсистема обучения и тонкой настройки. Она предназначена для автоматической адаптации модели к новым данным и включает в себя набор модулей, представленных на рис. 3.



Рис. 3. Компоненты подсистемы обучения и тонкой настройки нейро-нечетких компонент (ННК)

Модуль подготовки, валидации и нормализации данных, отвечающий за формирование обучающих выборок из сырьевых производственных данных. В рамках данного модуля происходит сбор информации о функционировании сложной системы за заданный период из различных источников (датчики оборудования, системы учёта качества, журналы операторов и т.д.). Далее данные проходят валидацию – проверку на соответствие критериям полноты и корректности (например, отсеивание аномальных или противоречивых записей). После очистки данные нормализуются – приводятся к единому масштабу, чтобы ни один параметр не доминировал из-за размерности. Также этот модуль разбивает данные на обучающую и тестовую выборки для последующей оценки качества обучения модели. Результатом работы модуля являются эталонные размеченные данные – множество векторов параметров с указанием исхода (успех/брак), разнесённых по тематическим классам факторов: оборудование, материалы, персонал, продукция, результат операции.

$$I = \left\{ \begin{array}{l} \{eq_1, eq_2, \dots, eq_n\} \\ \{mt_1, mt_2, \dots, mt_m\} \\ \{pr_1, pr_2, \dots, pr_l\} \\ \{st_1, st_2, \dots, st_{tk}\} \end{array} \right\},$$

где eq – характеристики оборудования, mt – параметры материалов, pr – свойства продукции, st – характеристики сотрудников.

Модуль фаззификации данных: преобразует количественные (числовые) параметры класса «результат операции» в качественные нечеткие оценки. Например, фактическое отклонение размера детали от номинала можно перевести в лингвистическую переменную «отклонение: малое/среднее/большое» с соответствующей функцией принадлежности, что необходимо для сравнения с результатом, получаемым от нейро-нечеткого компонента, и вычисления ошибки.

Модуль вычисления ошибки, предназначенный для сравнения нечеткого вывода нейро-нечеткого компонента с эталонным нечетким результатом и вычисления сигнала ошибки. Ошибка определяется, как разница (например, по некоторой метрике близости нечетких множеств) между предсказанным исходом операции и фактическим исходом (в нечетком представлении). Полученный сигнал ошибки используется для корректировки параметров нейро-нечеткого компонента – он поступает на алгоритм обучения (например, обратное распространение ошибки или его модификацию, адаптированную под нечеткие нейроны):

$$\tilde{O}_e \setminus \tilde{O} = \max\{\tilde{O}_e(x) - \tilde{O}(x), 0\}, x \in X.$$

Нейро-нечеткий компонент (ННК) представляет собой модель на основе описанного выше нечеткого нейрона Квана–Кэи. Он принимает на вход подготовленные четкие вектора параметров (разбитые по классам факторов: оборудование (*eq*), материалы (*mt*), персонал (*st*), параметры продукции (*pr*)) и выдает нечеткое множество исходов операции. Поскольку на производстве существуют разные типы технологических операций (сборка, сварка, окраска, проверка и т.п.), то для каждой категории операций настраивается свой экземпляр ННК с отдельным набором синаптических весов, наиболее подходящих для данного процесса. Выходы ННК показанные на рис. 4, интерпретируются в терминах трех классов:

- ♦ исход операции успешен – проведение данной технологической операции с заданными условиями не приведет к браку, результат соответствует стандартам качества;
- ♦ исход требует внимания – результат операции находится на грани допуска или есть повышенная неопределенность. Требуется принять решение о проведении дополнительных контрольных мероприятий (например, усиленный контроль качества, корректировка процесса), чтобы снизить вероятность негативного исхода;
- ♦ исход операции неудачен (брак) – с высокой вероятностью результат не соответствует требованиям (сильное отклонение параметров, явный дефект) либо имеется прямая индикация брака. В этом случае необходимо либо скорректировать процесс до выполнения операции, либо отбраковать изделие, либо запланировать его переделку после операции.

Такая градация выходов позволяет не просто бинарно решать задачу вида «брак/не брак», а учитывать промежуточное состояние, когда система не уверена или качество под вопросом. Это соответствует принципам риск-ориентированного подхода к качеству: если есть риск брака – принять меры превентивно.

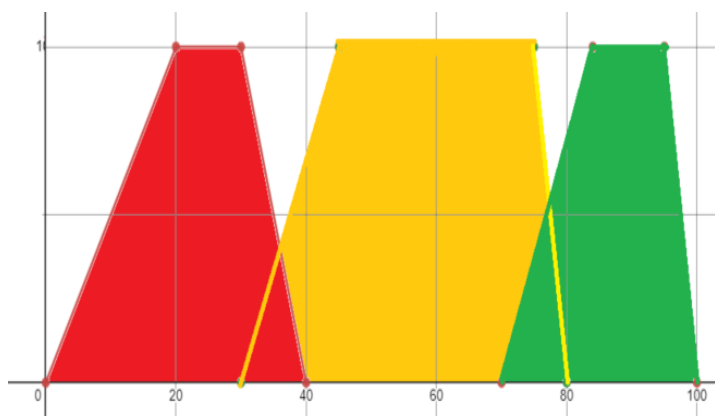


Рис. 4. Представление выходов нейро-нечеткого компонента

Модуль выбора конфигурации ННК: необходим в случае, если технологическая операция комплексная и задействует сразу несколько ресурсов (нескольких рабочих, единиц оборудования или материалов одновременно). Тогда операцию целесообразно

разбить на несколько более простых подопераций, каждая из которых оценивается своим нейро-нечетким компонентом. Модуль выбора решает, какую конфигурацию ННК применять для данной операции, исходя из её типа и задействованных ресурсов. По сути, это правило диспетчеризации: например, если операция «сборка» – использовать модель А, если «пайка» – модель В, если комбинированная – последовательно две модели и т.д.

В рамках подсистемы обучения предусмотрены два основных режима обучения нейро-нечеткого компонента (ННК):

- ♦ **полное обучение:** глобальное переобучение модели на всем массиве накопленных исторических данных за определенный период. Этот режим обеспечивает максимальную точность, но требователен к вычислительным ресурсам и времени, поэтому проводится относительно редко (например, при внедрении системы или при существенных изменениях процесса). Полное обучение позволяет ННК заново построить зависимости с нуля, учитывая весь накопленный опыт;

- ♦ **тонкая настройка:** периодическое дообучение модели на новых поступающих данных без полного переобучения. Используя свежие данные (например, последние смены или дни производства), ННК немного подстраивается под актуальные условия. Это быстрее и требует меньше ресурсов, позволяя поддерживать модель «в тонусе». Тонкую настройку целесообразно проводить в периоды низкой загрузки системы (например, во время ночных перерывов или технических окон), чтобы не влиять на производственный процесс. Она обеспечивает адаптацию к дрейфу процесса или медленным изменениям (например, постепенному износу оборудования, сезонным изменениям свойств материалов и т.д.).

Комбинация этих режимов обучения позволяет поддерживать высокую точность прогноза брака: редкие полные обучения дают глобальную корректировку модели, а регулярные тонкие настройки учитывают текущие изменения. Вся обучающая подсистема функционирует автономно, обеспечивая непрерывное улучшение интеллектуального компонента контроля качества по мере накопления данных.

Пример реализации на основе сетей Петри. Для демонстрации возможностей нейро-нечеткого подхода в управлении производством был разработан пример моделирования производственного процесса с использованием темпоральной нейро-нечеткой сети Петри (ТННСП). Сети Петри широко применяются для моделирования дискретных событийных систем, включая производственные процессы, благодаря наглядности и строгому математическому аппарату. В рассматриваемой модели ТННСП комбинируется с нечеткой логикой и временными параметрами.

Классические сети Петри описывают структуру процесса, но не учитывают явно время. Расширение до темпоральной сети Петри позволяет задавать временные задержки и интервалы для переходов и состояний, тем самым более точно отражая динамику реального производства. Данные особенности важны для планирования – например, учёт длительности операций, ожидания материалов, простоя оборудования. Модель с временем способна предсказывать поведение системы на шаг вперед, выявляя потенциальные конфликты или задержки.

Применение элементов с нечеткой логикой в узлах сети Петри позволяет оперировать с неточными понятиями (например, «высокая загрузка», «небольшая задержка») и неполной информацией. В реальном производстве часто данные бывают неточными или экспертными (оценочными). Нечеткая сеть Петри интегрирует правила вида «если состояние близко к перегрузке, то увеличить интервал обслуживания оборудования» и т.п., повышая реалистичность модели.

Нейро-нечеткая составляющая означает, что веса переходов и функции принадлежности в сети Петри могут настраиваться на основе данных или экспертного опыта, повышая адаптивность модели в целом, что делает её живым инструментом, способным эволюционировать вместе с системой.

Несмотря на сложность внутреннего устройства (нейросети, нечеткая логика, временные реляции), итоговая модель представима в виде наглядного графа, понятного инженерам, что облегчает восприятие и верификацию модели. Специалист по производству может интерпретировать элементы сети Петри, соотнеся их с этапами технологического процесса.

Подход на основе ТННСП является универсальным и может быть адаптирован и к другим предметным областям – управление транспортными системами, медицинскими процессами, финансами и т.д., везде, где есть процессы с событиями во времени и неопределённостью. В нашем случае, мы применяем её для задач планирования производства сложных технических объектов.

Формально темпоральная нейро-нечеткая сеть Петри определяется следующим образом.

$$TNPN = (PN, C, \alpha, f, B, R, Pl, m_0),$$

Она включает:

- ◆ классическую сеть Петри $PN = (P, T, F)$ – множество позиций (P) и переходов (T) с отношениями (F) между ними;
- ◆ множество времён дуг $C = \{C_1, C_2, \dots, C_{nc}\}$ – каждому ребру могут сопоставляться интервалы времени, задающие задержку или временное окно срабатывания;
- ◆ вектор нечетких порогов $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_{nc}\}$ – пороги срабатывания переходов в нечетких условиях;
- ◆ функции принадлежности $f = \{f_1, f_2, \dots, f_{nc}\}$ – определяют степень выполнения нечетких условий для переходов;
- ◆ множество $B = \{b, bi, m, mi, o, oi, d, di, s, si, f, fi, e\}$ – базовые интервальные отношения логики Аллена для описания отношений во времени между событиями (предшествует, следует, совпадает, длится и т.д.) [15];
- ◆ множество ресурсов $R = \{R_1, R_2, \dots, R_{nr}\}$, учитываемых в модели (оборудование, люди, оснастка, ограниченные ресурсы);
- ◆ производственный план Pl , задающий последовательности операций, партии изделий, сроки и пр.;
- ◆ начальную маркировку m_0 – вектор, задающий начальное состояние (расположение маркеров) с учетом нечеткой наличия маркеров в позициях.

На основе этой модели можно описывать, как выполняются операции, как перетекают маркеры (например, маркер – это деталь на определённой стадии, переход – выполнение операции, позиции – состояние детали: ожидание, обработка, проверка). Использование темпоральных нейро-нечетких сетей Петри позволяет учесть временные аспекты и неопределённость информации, достигая более высокой точности прогнозирования, чем традиционные методы, которые зачастую рассматривают усреднённые или детерминированные процессы [16], что особенно актуально для сложного планирования: например, модель способна показать, что если одна операция задержится (с вероятностью, характеризующей нечетко), то последующие сдвинутся и возможен простой оборудования, влияющий на весь план.

Разработанные программно-инструментальные средства. Для практического применения описанных методов разработано специализированное программное обеспечение. Оно интегрируется в инфраструктуру предприятия и обеспечивает сбор данных, обучение модели и поддержку принятия решений в режиме, близком к реальному времени [17]. Основные возможности созданного программно-инструментального комплекса включают:

- ◆ формирование онтологической модели сложной системы. Пользователь или эксперт может задать онтологию производства – классы информационных сущностей (оборудование, операции, материалы, дефекты и т.п.) и связи между ними. Такая модель знаний служит основой для настройки нейро-нечеткой системы, позволяя ей учитывать структуру предметной области (например, что определённый тип станка связан с определёнными видами дефектов и операциями);

- ♦ валидация внешних и аналитических данных нейросетевым супервизором. Встроенный модуль на базе нейросети контролирует качество поступающих данных. Он может отслеживать, например, аномалии (всплески показаний датчиков, сбои в передаче данных) и сигнализировать о них, тем самым обеспечивая надежность входной информации для нашей системы;

♦ средства моделирования и прогнозирования на основе ТННСП. Реализован графический конструктор, экранная форма изображена на рисунке 5, позволяющий специалистам строить модели процессов в виде темпоральных нечетких сетей Петри, не погружаясь в кодирование алгоритмов. Визуально смоделировав процесс и задав начальные условия, пользователь может запустить имитацию и получить прогноз – например, выявить узкие места или оценить риск брака на каждом этапе. Отмечено, что смоделированные таким образом траектории производственного процесса на ~80% совпадают с решениями, принимаемыми экспертами вручну на основе опыта (что свидетельствует о достаточной адекватности модели);

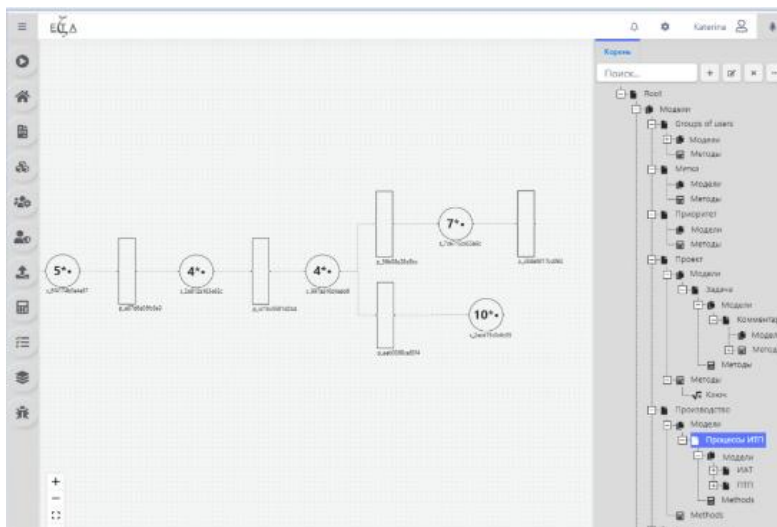


Рис. 5. Экранная форма разработанного инструмента

Во время внедрения на предприятии было показано, что подобная модель может успешно поддерживать задачу календарного планирования производства. Построенное программное приложение позволило проектировать информационно-технологические процессы с помощью конструктора моделей на базе ТННСП. Выяснилось, что траектории процессов, сгенерированные моделью, в 80% случаев совпадают с планами, составленными экспертами-технологами вручную (расхождения касались в основном нештатных ситуаций, которые модель пока не учитывала детально). Такой результат демонстрирует перспективность предложенного подхода для автоматизации планирования и прогнозирования в производственных системах.

Заключение. Проведенное исследование показало эффективность использования нейро-нечетких компонентов для управления качеством и прогнозирования производственного брака на мелкосерийных предприятиях. Предложенные модели и алгоритмы позволяют на ранней стадии обнаруживать потенциальные дефекты и принимать упреждающие меры, тем самым существенно сокращая издержки на бракованную продукцию и переделки.

В работе обосновано, что нейро-нечеткие системы способны объединить преимущества адаптивного обучения и логической интерпретируемости, что особенно ценно в условиях сложных и вариативных процессов. Разработанный нейро-нечеткий компонент на основе нечеткого нейрона Квана–Кэи успешно справляется с задачей классификации исходов технологических операций, а подсистема его обучения обеспечивает постоянную адаптацию к изменениям в процессе. Интеграция временных и нечетких моделей (темпоральных нечетких сетей Петри) продемонстрировала возможность более точного планирования и имитации производственных процессов, учитывающих неопределенность и динамику.

Практическая реализация в виде программного комплекса подтвердила применимость метода: интеллектуальная система качества может работать в реальном времени параллельно с производством, не нарушая его, и генерировать ценные рекомендации. Таким образом, цели исследования достигнуты – показано, что применение нейро-нечетких методов действительно повышает эффективность управления качеством на мелкосерийных производствах.

Перспективы дальнейших исследований видятся в развитии специализированных нейро-нечетких алгоритмов именно для производственных задач. Например, создание гибридных моделей, объединяющих нейро-нечеткий подход с методами статистического контроля (SPC) для учета как данных прошлого, так и онлайн-сигналов. Интересным направлением является интеграция предложенной системы с концепцией Цифрового двойника производства, когда нейро-нечеткая модель станет ядром цифрового двойника, постоянно обучающегося на данных и управляющего качеством в замкнутом цикле. Также планируется расширить базу знаний системы за счёт добавления правил экспертной системы, чтобы улучшить интерпретируемость рекомендаций для операторов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Perfecting the 1:10:100 Rule in Data Quality // Medium.com: электронный блог. 17.03.2025. – URL: <https://medium.com/grepsr-blog/perfecting-the-1-10-100-rule-in-data-quality-a1f31143f40b> (дата обращения: 17.03.2025).
2. Misnik A., Borisov V. Compositional neuro network modeling of complex technical systems // *Neurocomputers: Development, Application*. – 2016. – No. 7. – P. 39-46.
3. Таненбаум Э. Распределенные системы: принципы и парадигмы. – СПб.: Питер, 2003. – 877 с.
4. Haykin S. *Neural networks: A comprehensive foundation*. – 2nd ed. – New Jersey: Prentice Hall International, Inc, 1999. – 1103 p.
5. Hernández S., Sáez D., Mery D. Neuro-Fuzzy Method for Automated Defect Detection in Aluminium Castings / In: Campilho, A., Kamel, M. (eds.) // *Image Analysis and Recognition. ICIAR 2004. Lecture Notes in Computer Science*. – Vol. 3212. – Springer, Berlin, Heidelberg, 2004.
6. Орлов А.И. Теория принятия решений. – М.: Экзамен, 2005. – 656 с.
7. Тарасик В.П. Математическое моделирование технических систем: учебник для вузов. – Мн.: ДизайнПРО, 2004. – 640 с.

8. Марка Д., МакГоуэн К. Методология структурного анализа и проектирования. – М.: МетаТехнология, 1993. – 240 с.
9. Bobryakov A.V., Krutalevich S.K., Misnik A.E., Prokopenko S.A. Modeling of Industrial and Technological Processes in Complex Systems Based on NeuroFuzzy Petri Nets // *Journal of Physics: Conference Series*, Volume 2096, International Conference on Automatics and Energy (ICAE 2021) 7-8 October 2021, Vladivostok, Russia.
10. Tax N., Verenich I., La Rosa M., Dumas M. Predictive Business Process Monitoring with LSTM neural networks // *Proceedings of the International Conference on Advanced Information Systems Engineering (CAiSE)*. – P. 477-492. arXiv:1612.02130 Freely accessible. – doi: 10.1007/978-3-319-59536-8_30.
11. Попов Э.В. Экспертные системы. – М.: Наука, 1987. – 288 с.
12. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. – СПб.: БХВ-Петербург, 2003.
13. Ash T. Dynamic Node Creation in Back-Propagation Networks // *II Connection Science*. – 1989. – Vol. 1.
14. Kwan H., Cai L. A fuzzy neural network and its application to pattern recognition // *IEEE Transactions on Fuzzy Systems*. – 1994. – Vol. 2, No. 3. – P. 185-193. – DOI: 10.1109/91.298447.
15. Allen J.F. Maintaining knowledge about temporal intervals // *Communications of the ACM*. – 1983. – Vol. 26, No. 11. – P. 832-843.
16. Wu J. and Yan S. Reliability Evaluation for Mechanical Systems by Petri Nets // *Petri Nets in Science and Engineering*. – 2018.
17. Balbiani P., Boudou J., Diéguez M., Fernández-Duque D. Intuitionistic Linear Temporal Logics // *ACM Transactions on Computational Logic*. – 2019. – 21. – P. 1-32.
18. Vajnilovich Yu., Zaharchenkov K. and Zakharova A. Integrated System Approach to Improving the Efficiency of IT Projects Management Based on Evolutionary Modeling // *Computing in Physics and Technology*. – P. 309-314.
19. Borisov V.V., Misnik A.E. Ontological Engineering of Interrelated Processes in Complex Cyber-Physical Systems / In: Kovalev, S., Sukhanov, A., Akperov, I., Ozdemir, S. (eds) // *Proceedings of the Sixth International Scientific Conference “Intelligent Information Technologies for Industry” (IITI’22)*. IITI 2022. Lecture Notes in Networks and Systems. – Vol. 566. – Springer, Cham, 2023. – https://doi.org/10.1007/978-3-031-19620-1_39.
20. Karabach A. Information integration systems based on semantic technologies // *Science, technology and education*. – 2014. – Vol. 2 (2).

REFERENCES

1. Perfecting the 1:10:100 Rule in Data Quality // Medium.com: электронный блог. 17.03.2025. Available at: <https://medium.com/grepsr-blog/perfecting-the-1-10-100-rule-in-data-quality-a1f31143f40b> (accessed 17 March 2025).
2. Misnik A., Borisov V. Compositional neuro network modeling of complex technical systems, *Neurocomputers: Development, Application*, 2016, No. 7, pp. 39-46.
3. Tanenbaum E. Распределенные системы: принципы и парадигм [Distributed systems: principles and paradigms]. Saint Petersburg: Piter, 2003, 877 p.
4. Haykin S. Neural networks: A comprehensive foundation. 2nd ed. New Jersey: Prentice Hall International, Inc, 1999, 1103 p.
5. Hernández S., Sáez D., Mery D. Neuro-Fuzzy Method for Automated Defect Detection in Aluminium Castings, In: Campilho, A., Kamel, M. (eds.), *Image Analysis and Recognition. ICIAR 2004. Lecture Notes in Computer Science*, Vol. 3212. Springer, Berlin, Heidelberg, 2004.
6. Orlov A.I. Теория принятия решений [Decision theory]. Moscow: Ekzamen, 2005, 656 p.
7. Tarasik V.P. Математическое моделирование технических систем: учебник для вузов [Mathematical modeling of technical systems: a textbook for universities]. Minsk: DizaynPRO, 2004, 640 p.
8. Марка Д., МакГоуэн К. Методология структурного анализа и проектирования [Methodology of structural analysis and design]. Moscow: MetaTekhnologiya, 1993, 240 p.
9. Bobryakov A.V., Krutalevich S.K., Misnik A.E., Prokopenko S.A. Modeling of Industrial and Technological Processes in Complex Systems Based on NeuroFuzzy Petri Nets, *Journal of Physics: Conference Series*, Volume 2096, International Conference on Automatics and Energy (ICAE 2021) 7-8 October 2021, Vladivostok, Russia.
10. Tax N., Verenich I., La Rosa M., Dumas M. Predictive Business Process Monitoring with LSTM neural networks, *Proceedings of the International Conference on Advanced Information Systems Engineering (CAiSE)*, pp. 477-492. arXiv:1612.02130 Freely accessible. doi: 10.1007/978-3-319-59536-8_30.
11. Попов Э.В. Экспертные системы [Expert systems]. Moscow: Nauka, 1987, 288 p.

12. Leonenkov A.V. Nechetkoe modelirovanie v srede MATLAB i fuzzyTECH [Fuzzy modeling in MATLAB and FuzzyTECH]. Saint Petersburg: BKhV-Peterburg, 2003.
13. Ash T. Dynamic Node Creation in Back-Propagation Networks, *II Connection Science*, 1989, Vol. 1.
14. Kwan H., Cai L. A fuzzy neural network and its application to pattern recognition // *IEEE Transactions on Fuzzy Systems*. – 1994. – Vol. 2, No. 3. – P. 185-193. – DOI: 10.1109/91.298447.
15. Allen J.F. Maintaining knowledge about temporal intervals, *Communications of the ACM*, 1983, Vol. 26, No. 11, pp. 832-843.
16. Wu J. and Yan S. Reliability Evaluation for Mechanical Systems by Petri Nets, *Petri Nets in Science and Engineering*, 2018.
17. Balbiani P., Boudou J., Diéguez M., Fernández-Duque D. Intuitionistic Linear Temporal Logics, *ACM Transactions on Computational Logic*, 2019, 21, pp. 1-32.
18. Vajnilovich Yu., Zaharchenkov K. and Zakharova A. Integrated System Approach to Improving the Efficiency of IT Projects Management Based on Evolutionary Modeling, *Computing in Physics and Technology*, pp. 309-314.
19. Borisov V.V., Misnik A.E. Ontological Engineering of Interrelated Processes in Complex Cyber-Physical Systems, In: Kovalev, S., Sukhanov, A., Akperov, I., Ozdemir, S. (eds), *Proceedings of the Sixth International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'22). IITI 2022. Lecture Notes in Networks and Systems*, Vol. 566. Springer, Cham, 2023. Available at: https://doi.org/10.1007/978-3-031-19620-1_39.
20. Karabach A. Information integration systems based on semantic technologies, *Science, technology and education*, 2014, Vol. 2 (2).

Прокопенко Сергей Александрович – Белорусско-Российский университет; e-mail: puss95@yandex.by; г. Могилев, Беларусь; тел.: +375447535081; +79679872264; ассистент; ассистент НИУ «МЭИ».

Бобряков Александр Владимирович – НИУ «МЭИ»; e-mail: avbob@mail.ru; г. Москва, Россия; тел.: +74953627407; д.т.н.; зав. кафедрой УИТ.

Prokopenko Sergey Aleksandrovich – Belarusian-Russian University; e-mail: puss95@yandex.by; Mogilev, Belarus; phones: +375447535081; +79679872264; assistant; assistant of National Research University "Moscow Power Engineering Institute".

Bobryakov Alexander Vladimirovich – National Research University "Moscow Power Engineering Institute"; e-mail: avbob@mail.ru; Moscow, Russia; phone: +74953627407; dr. of eng. sc.; head of the Department of Information Technologies;

УДК 004.932

DOI 10.18522/2311-3103-2026-1-284-295

В.Е. Бондарева, Т.С. Черноморова, А.В. Кривцун, А. Абдулкарем

АЛГОРИТМ РАСПОЗНАВАНИЯ ЯЗЫКА ЖЕСТОВ С ИСПОЛЬЗОВАНИЕМ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ

Рассматривается задача автоматического распознавания русского жестового языка (РЖЯ) на основе методов компьютерного зрения и глубокого обучения. Актуальность исследования обусловлена устойчивым ростом численности лиц с нарушениями слуха: по данным Всемирной организации здравоохранения, в настоящее время их количество составляет около 70 млн человек, и, согласно прогнозам, может увеличиться до 630 млн к 2035 году. Разработка эффективных алгоритмов распознавания жестов является важным направлением создания систем бесконтактного человеко-машинного взаимодействия, направленных на повышение доступности информационных технологий и улучшение качества жизни людей с ограниченными возможностями слуха. Целью работы является разработка и экспериментальная апробация алгоритма распознавания букв русского жестового языка в видеопотоке в реальном времени с использованием сверточной нейронной сети. В ходе исследования сформирован специализированный датасет, включающий 430 изображений жестов, соответствующих буквам алфавита РЖЯ, представленных с различных ракурсов и при изменяющихся условиях освещения. Для реализации модели использованы библиотеки TensorFlow и Keras, а интеграция с видеопотоком выполнена с применением средств OpenCV и системы отслеживания положения руки на основе маркеров. В результате проведенного обучения и тестирования предложенная модель продемонстрировала точность распознавания на уровне 99

% на тестовой выборке. Проведён сравнительный анализ с классическими методами машинного обучения, показавший преимущество сверточной нейронной сети по точности классификации и устойчивости к внешним помехам. Полученные результаты подтверждают эффективность предложенного подхода и его применимость для создания систем реального времени, предназначенных для коммуникационных, образовательных и реабилитационных приложений, а также для разработки интерфейсов человеко-машинного взаимодействия.

Язык жестов; сверточная нейронная сеть; машинное обучение.

V.E. Bondareva, T.S. Chernomorova, A.V. Krivtsun, A. Abdulkarem

ALGORITHM FOR SIGN LANGUAGE RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK

This paper addresses the problem of automatic recognition of Russian Sign Language (RSL) using computer vision and deep learning methods. The relevance of the study is driven by a steady increase in the number of people with hearing impairments: according to the World Health Organization, there are currently about 70 million deaf and hard-of-hearing individuals worldwide, and this number is projected to reach 630 million by 2035. The development of effective gesture recognition algorithms is an important direction for creating contactless human-machine interaction systems aimed at improving accessibility of information technologies and enhancing the quality of life for people with hearing disabilities. The aim of the study is to develop and experimentally validate an algorithm for real-time recognition of Russian Sign Language alphabet gestures in a video stream using a convolutional neural network. A specialized dataset was created, consisting of 430 images of hand gestures corresponding to the letters of the RSL alphabet, captured from different angles and under varying lighting conditions. The model was implemented using TensorFlow and Keras libraries, while integration with the video stream was performed using OpenCV and a marker-based hand tracking system. As a result of training and testing, the proposed model achieved a recognition accuracy of 99% on the test dataset. A comparative analysis with classical machine learning methods demonstrated the superiority of the convolutional neural network in terms of classification accuracy and robustness to external noise. The obtained results confirm the effectiveness of the proposed approach and its applicability for real-time systems intended for communication, educational, and rehabilitation applications, as well as for the development of advanced human-machine interaction interfaces.

Sign language; convolutional neural network; machine learning.

Введение. К концу 2023 года по всему миру насчитывалась около 70 миллионов глухонемых людей. К сожалению, эта цифра продолжит расти и в будущем, так как потеря слуха может произойти не только из-за болезни, но и из-за внешних факторов, как громкие и постоянные звуки на протяжении длительного времени. По предварительным подсчётам количество глухонемых может возрасти до 630 миллионов человек к 2035 и до 900 миллионов человек к 2050 году [1]. Люди с нарушениями слуха общаются на языке жестов. Язык жестов состоит из множества символов, которые для каждой страны разные [2].

Компьютерное зрение помогает распознавать жесты, но для этого необходимо разработать и обучить специальные алгоритмы и нейросети. Для компьютерного зрения руки являются сложным объектом. На руках нет "активных" точек, подобно глазам и рту на лице, что затрудняет естественное восприятие языка жестов [3]. Для реализации подобных систем существует два основных подхода:

- ♦ **Маркерный подход.** В этом подходе на руки человека наносятся маркеры, которые могут быть оптическими, магнитными или акустическими. Камера, установленная на устройстве, считывает эти маркеры и определяет жесты [4].

- ♦ **Безмаркерный подход.** Этот подход использует алгоритмы компьютерного зрения для анализа движений рук и пальцев без использования дополнительных маркеров. Нейросеть анализирует жесты, используя алгоритмы компьютерного зрения и методы обработки изображений. Программа получает изображение с камеры, обрабатывает его и определяет различные параметры, такие как положение, ориентация и форма жеста. Затем нейросеть классифицирует жест на основе полученных данных и сравнивает его с известными жестами, хранящимися в её базе знаний. В результате модель может определить, что означает данный жест [5].

Сравнительные исследования показывают преимущества сверточных нейронных сетей (CNN) над классическими искусственными нейронными сетями (ИНС) за счёт эффективного выявления пространственных и временных признаков. Искусственная нейронная сеть (ИНС) – классическая сеть, в которой каждый нейрон связан с каждым на следующем слое; эффективно работает с векторизованными признаками, но не учитывает пространственную структуру изображений. Сверточная нейронная сеть (CNN) – специализированная архитектура для работы с изображениями; использует свертки для выявления локальных признаков, что обеспечивает высокую точность распознавания при анализе пространственно-временных паттернов жестов [6]. CNN обладают уникальной архитектурой, которая позволяет им эффективно выявлять пространственные признаки в изображениях. Это достигается за счет использования сверток, которые позволяют обнаруживать локальные паттерны и текстуры, что особенно важно для анализа движений рук и пальцев. Благодаря этому, CNN могут точно распознавать формы и положение объектов, что критично для задач, связанных с жестами и манипуляциями. [7]. Для анализа движений рук важна не только пространственная информация, но и временная динамика. Хотя традиционные CNN не предназначены для обработки последовательностей данных, их можно комбинировать с другими архитектурами, такими как долгосрочная краткосрочная память (LSTM) или временные сверточные сети (TCN). Это позволяет эффективно захватывать изменения во времени и улучшать точность распознавания движений [8]. CNN также показывают высокую устойчивость к шуму в данных. Это особенно важно в реальных условиях, где изображения могут быть искажены или содержать лишнюю информацию. За счет своей способности к обобщению и извлечению ключевых признаков из сложных входных данных, CNN обеспечивают надежность в распознавании жестов даже при наличии помех. Сверточные нейронные сети легко масштабируются для работы с различными размерами входных данных и могут быть адаптированы под конкретные задачи. Это делает их универсальным инструментом для разработки систем анализа движений рук и пальцев, которые могут быть использованы в различных приложениях – от виртуальной реальности до реабилитации. Кроме того, сверточные сети могут быть более простыми и эффективными в сравнении с полносвязными сетями, что делает их более подходящими для реализации на устройствах с ограниченными ресурсами.

Современные исследования показывают актуальность выбранной темы исследования. Исследование Байхана и соавторов (2024) предложило модифицированную модель CNNsa-LSTM, которая объединяет сверточную нейронную сеть (CNN) для извлечения пространственных признаков, механизм самовнимания (Self-Attention, SA) для выделения ключевой информации и рекуррентную LSTM для моделирования временной динамики жестов. Такая архитектура позволила достичь высокой точности – около 96,4% – при классификации жестов языка глухонемых на видео. Использование BLSTM (двухнаправленной LSTM) улучшает устойчивость и качество распознавания продолжительных сигналов, важных для перевода сложных жестов [9]. Кумар с коллегами (2022) реализовали систему Markerless hand gesture recognition (Безмаркерное распознавание с глубоким обучением), которая использует методы глубокого обучения и сверточные сети для сегментации и анализа движений рук в реальном времени. Их алгоритм показал высокую точность классификации даже в сложных условиях освещения и фона, что подтверждает потенциал безмаркерных методов в практическом применении для распознавания языка жестов [10]. Работа Амёра и соавторов (2025) представила сложную архитектуру пространственно-временной модели с генеративной сетью, сочетающую deep LSTM autoencoder и рекуррентные сети, что позволило не только классифицировать жесты, но и восстанавливать последовательности для генерации дополнительных обучающих данных. Это увеличивает обобщающую способность модели и устойчивость к шуму, что критично для реального применения в распознавании жестов [11]. Аль-Мудави и коллеги (2024) разработали комплексную систему для распознавания жестов в реальных условиях, применив несколько модулей, включая выделение скелета руки, глубокую сверточную сеть для детекции и алгоритм оптимизации на основе народного обучения. Для классификации жестов использовалась 1D CNN, что позволило достичь высокой точно-

сти на видео с динамическими движениями рук, обеспечивая устойчивость результатов в шумных видеопотоках [12]. Пател и соавторы (2023) акцентировали внимание на оптимизации CNN-моделей для работы на мобильных и встроенных устройствах без значительной потери точности – около 90% точности удалось сохранить при уменьшении вычислительных затрат, что крайне важно для внедрения систем распознавания языка жестов в повседневные гаджеты [13].

CNN способны выделять локальные паттерны и текстуры на изображениях, что важно для анализа движений рук. Они также более устойчивы к шуму и легко масштабируются для работы с различными размерами входных данных. Современные исследования показывают высокую точность CNN в сочетании с рекуррентными архитектурами (LSTM, BLSTM, TCN) для анализа динамических жестов [7–13].

Постановка задачи.

Научная проблема:

Современные системы распознавания языка жестов сталкиваются с рядом трудностей: высокая вариативность исполнения жестов, различие ракурсов, освещения, частичное перекрытие пальцев, шумы видеопотока. Существующие методы (классические ИНС, логистическая регрессия, гибридные CNN-LSTM) либо не обеспечивают достаточной точности, либо требуют высокой вычислительной мощности, что затрудняет их применение в реальном времени.

Объект исследования:

Объект исследования – символы русского жестового языка (РЖЯ), формируемые и отображаемые посредством движений рук и пальцев человека.

Предмет исследования:

Алгоритмы и методы машинного обучения, способные анализировать пространственные и временные признаки видеоданных для распознавания жестов. В частности, сверточные нейронные сети (CNN) как инструмент выделения локальных признаков в изображениях и их комбинация с маркерным и безмаркерным подходами.

Цель исследования:

Создание высокоточной системы распознавания языка жестов на базе сверточной нейронной сети, обеспечивающей работу в реальном времени, с возможностью интеграции в системы человеко-машинного взаимодействия.

Задачи исследования:

Для достижения цели необходимо решить следующие задачи:

- ♦ Создание датасета жестов РЖЯ: сбор видеок кадров, аннотирование изображений и разметка символов с различными ракурсами и освещением.
- ♦ Разработка модели сверточной нейронной сети: выбор архитектуры, слоев, функций активации, методов нормализации и регуляризации для эффективного выделения пространственных признаков жестов.
- ♦ Обучение и тестирование модели: подбор оптимальных параметров обучения, контроль за переобучением, оценка точности распознавания на тестовой выборке.
- ♦ Интеграция отслеживания руки с видеопотоком: использование маркеров и алгоритмов компьютерного зрения для устойчивого определения положения кисти и пальцев в реальном времени.
- ♦ Оценка эффективности алгоритма: сравнение с другими методами (ИНС, логистическая регрессия, CNN-LSTM) по критериям точности, скорости и устойчивости к шумам.

Научная гипотеза:

Использование сверточной нейронной сети с интеграцией маркерного отслеживания руки и специализированной архитектуры слоев позволит достичь более высокой точности распознавания жестов ($\geq 95\%$) и обеспечить работу алгоритма в реальном времени, превосходя классические методы и гибридные решения по сочетанию точности и скорости.

Методика исследования. Датасет включал 430 изображений букв РЖЯ (10–15 кадров на каждый жест), обработанных с использованием LabelImg. Данные были разделены на обучающую (80%) и тестовую (20%) выборки.

Модель реализована в Python с использованием Keras и TensorFlow. Архитектура CNN включала:

- ♦ 3 слоя Conv2D с 64 фильтрами, 2 слоя Conv2D с 128 фильтрами (ядро 3×3), ReLU активация.
- ♦ Слои Dropout для предотвращения переобучения.
- ♦ BatchNormalization для нормализации входных данных.
- ♦ MaxPooling для снижения размерности карт признаков.
- ♦ Flatten и Dense слои для классификации жестов.

При подаче изображения или видео на вход система определяет положение руки. После этого процесс разделяется на две ветки:

1) Работа с визуальной частью. Система определяет положение каждого пальца руки, после чего накладывает на неё маркеры. Чтобы процесс не повторялся каждый раз, программа отслеживает руку и обновляет маркеры уже найденных пальцев.

2) Работа с нейросетью. После нахождения руки её изображение отправляется в нейронную сеть, где производится сопоставление с уже известной модели информацией. Если жест совпадает, то выдаётся перевод нужного жеста. Если жест отсутствует в системе, то подбирается наиболее похожий символ.

Пример изображений для буквы С представлен на рис. 1.



Рис. 1. Набор изображений для буквы С

Для разметки датасета использовался инструмент *LabelImg* – графический инструмент для аннотирования изображений, созданный на языке *Python*. [14].

После разметки данных все изображения были разбиты на два датасета: обучающая и тестовая. Данные были поделены 80% на 20%, то есть 8 изображений каждого жеста в обучающей выборке (344 изображения в сумме) и 2 изображения каждого жеста в тестовой выборке (86 изображений).

Для разработки был выбран язык программирования *Python*. Использовались библиотеки *Keras*, *TensorFlow* и *Scikit-learn*, которые применяются в области машинного обучения. Отдельного внимания заслуживает библиотека *OpenCV* – это библиотека компьютерного зрения с открытым исходным кодом для анализа, классификации и обработки изображений, широко используемая в таких языках, как *C*, *C++*, *Python* и *Java* [15].

Для распознавания жестовых символов использование данной библиотеки является важным. С её помощью создан датасет, который будет записан в режиме реального времени. После обучения модели *OpenCV* сможет определять жесты и предлагать перевод непосредственно в видео.

Модель алгоритма. Для обучения модели распознавания языка жестов была выбрана свёрточная нейронная сеть (*CNN*). Для сравнения применялись три алгоритма, используемые для работы с изображениями: свёрточная нейронная сеть, логистическая регрессия и искусственная нейронная сеть. Исходя из результатов тестирования на тестовом датасете с жестами цифр именно *CNN* подходит лучше всего.

Разработка модели происходила в *Jupyter Notebook*. Обученная модель сохраняется в формате *H5* для дальнейшего использования в видеопотоке. Сам формат *H5* (*Hierarchical Data Format version*) предназначен для хранения и организации больших объёмов данных. Он сохраняет веса модели, обрабатывает сложные иерархические структуры и поддерживает метаданные [16].

Модель в формате *H5* выгружается в скрипт с расширением «*.py*», где происходит распознавание языка жестов в реальном времени из потока данных от видеокамеры. В первую очередь требуется обучить искусственную нейронную сеть распознавать буквы на изображениях и определять их принадлежность к алфавиту. Для этого необходимо разработать функцию и использовать встроенные методы библиотеки *Keras* для создания и обучения модели. Для усовершенствования разработанной модели внедряются дополнительные параметры следующим образом:

1) *model = Sequential()* – Создается объект модели типа *Sequential*, который позволяет последовательно добавлять слои. *Sequential* представляет собой последовательность слоёв, где каждый слой выполняет определённую операцию над входными данными [17]. Этот тип идеально подходит для решения задачи распознавания языка жестов, так как основным принципом алгоритма является классификация и обработка изображений.

2) *Conv2D* – это слой свертки в нейронных сетях, который применяется для обнаружения различных признаков в изображениях. Он используется для свертки входных данных и для извлечения признаков. Каждый фильтр скользит по изображению, применяя операцию свертки, и создает новое изображение (*feature map*) с выделенными признаками [18]. У *Conv2D* есть несколько параметров, включая количество фильтров (*filters*), размер ядра свертки (*kernel_size*), функцию активации (*activation*), способ заполнения (*padding*), шаг (*stride*) и другие. Количество фильтров определяет количество признаков, которые будут извлечены на каждом уровне свертки. Размер ядра свертки определяет область, по которой будет скользить фильтр. Функция активации применяется к выходным данным каждого нейрона. Параметры *padding* и *stride* контролируют размер выходного изображения и шаг, соответственно. Комбинация этих параметров определяет поведение *Conv2D* слоя в нейронной сети.

Для модели создавалось три слоя типа *Conv2D* с 64 фильтрами размером 3x3 и два слоя с 128 фильтрами размером 3x3 со следующими параметрами:

1) *padding = 'Same'* – специальная форма маскировки *padding*, основанная на кодировке последовательностей, установлена в «*Same*». Благодаря этому входные карты признаков будут иметь ту же размерность, что и выходные;

2) *activation = 'relu'* – функция активации *ReLU*;

3) *input_shape = (28,28,1)* – входная форма (28,28,1), что соответствует изображениям размером 28x28 пикселей с одним каналом (например, черно-белое изображение);

4) *Dropout()* - Добавляется слой *Dropout* с вероятностью исключения 0.1, что помогает предотвратить переобучение. В коде таких слоёв несколько с исключением 0.2 и 0.3;

5) *BatchNormalization()* – Добавляется слой *BatchNormalization* для нормализации входных данных каждого мини-пакета. Нормализация пакетов (*batch-normalization*) – метод, способствующий увеличению производительности и устойчивости искусственных нейронных сетей [14]. Основная идея этого метода заключается в том, что некоторые слои нейронной сети получают на вход данные, которые были предварительно обработаны таким образом, чтобы иметь нулевое математическое ожидание и единичную дисперсию;

6) *MaxPool2D(2,2)* – Добавляется слой *MaxPooling* с размером пула 2x2, который уменьшает размер карт признаков вдвое. Слой *Pooling* в сверточных нейронных сетях используется для уменьшения размерности карты признаков. Он имеет две разновидности: *max-pooling* и *average-pooling*. В данной нейросети используется *max-pooling* для максимального сокращения;

7) *Flatten()*– Карты признаков после последнего слоя *MaxPooling* выравниваются в одномерный вектор. *Flatten* позволяет преобразовать структурированные данные, такие как изображения, в линейный формат, который удобно использовать в нейронных сетях;

8) *Dense(128, activation='relu')* – Добавляется полносвязный слой с 128 нейронами и функцией активации *ReLU (Rectified Linear Unit)*. Это нелинейная функция активации, которая преобразует входное значение в значение от 0 до положительной бесконечности. Если входное значение меньше или равно нулю, то *ReLU* выдаёт ноль, в противном случае – входное значение.

Сама модель принимает вид, изображенный на рис. 2.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 28, 28, 64)	640
dropout (Dropout)	(None, 28, 28, 64)	0
batch_normalization (Batch Normalization)	(None, 28, 28, 64)	256
max_pooling2d (MaxPooling2D)	(None, 14, 14, 64)	0
dropout_1 (Dropout)	(None, 14, 14, 64)	0
conv2d_1 (Conv2D)	(None, 14, 14, 128)	73856
batch_normalization_1 (Batch Normalization)	(None, 14, 14, 128)	512
max_pooling2d_1 (MaxPooling2D)	(None, 7, 7, 128)	0
dropout_2 (Dropout)	(None, 7, 7, 128)	0
conv2d_2 (Conv2D)	(None, 7, 7, 128)	147584
batch_normalization_2 (Batch Normalization)	(None, 7, 7, 128)	512
max_pooling2d_2 (MaxPooling2D)	(None, 3, 3, 128)	0
flatten (Flatten)	(None, 1152)	0
dense (Dense)	(None, 128)	147584
dropout_3 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 25)	3225
Total params: 374169 (1.43 MB)		
Trainable params: 373529 (1.42 MB)		
Non-trainable params: 640 (2.50 KB)		

Рис. 2. Модель нейронной сети

Модель важно скомпилировать. Это позволяет точно настроить процесс обучения под конкретную задачу и обеспечивает возможность мониторинга и анализа прогресса обучения. Без компиляции модель не сможет корректно обучаться. На данном этапе заканчивается подготовка нейронной сети к её обучению. Указанные слои и параметры позволяют модели достичь высокой точности результатов предсказаний, а компиляция позволит избежать ошибок и переобучения [19].

Для подключения нейросети к видеокамере важно настроить отслеживание руки с помощью точек. Для этого модель была обучена в модуле *mediapipe*:

1) *mp_hands* – предварительно обученную модель для обнаружения и отслеживания рук.

2) *mp_drawing* – предоставляет методы для рисования на изображениях.

3) *mp_drawing_styles* – стили для рисования рук.

Затем создаётся экземпляр класса *Hands* с параметрами *static_image_mode=True* и *min_detection_confidence=0,3*. Эти параметры указывают, что модель будет работать в режиме статического изображения и что минимальная уверенность в обнаружении должна быть 0,3 или выше.

Далее запускается бесконечный цикл, в котором считывается кадр с камеры. Кадр преобразуется из стандартного цветового пространства *BGR* в *RGB*. После этого кадр обрабатывается методом *Hands*, чтобы обнаружить руки. Если метод обнаруживает руки, то для каждой обнаруженной руки рисуются маркеры на кадре. Этот метод принимает следующие аргументы:

- 1) кадр, на котором нужно нарисовать маркеры;
- 2) маркеры, которые были обнаружены методом;
- 3) соединения между маркерами, которые нужно нарисовать;
- 4) стиль для рисования маркеров;
- 5) стиль для рисования соединений.

Были выбраны цветные маркеры, каждый цвет отвечает за отдельный палец: бежевый – большой, фиолетовый – указательный, жёлтый – средний, зелёный – безымянный и синий – мизинец. На каждой фаланге есть маркер, который обозначает место сгиба пальца. Это позволяет отслеживать пальцы в разных ракурсах, рассчитывать расположение скрытых частей. Маркер руки представлен на рис. 3.

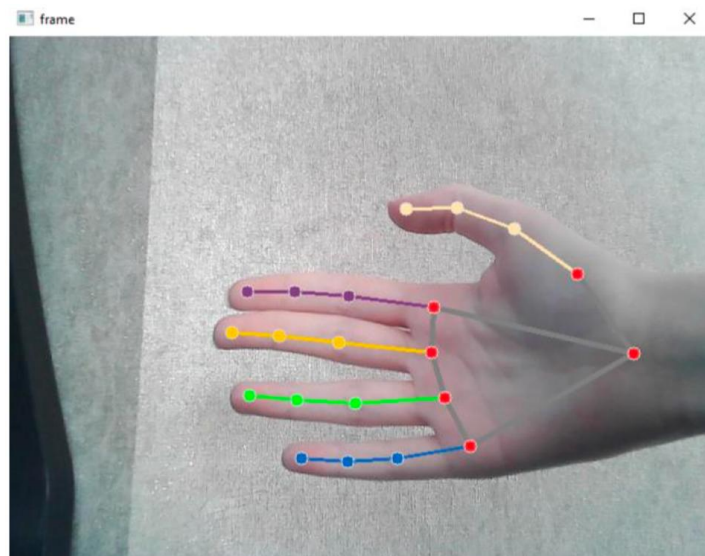


Рис. 3. Маркеры руки

После того как маркеры нарисованы, видео на экране отображается с помощью метода *imshow* библиотеки *OpenCV*. На экране в любом ракурсе отображаются маркеры на руке. Само предсказание будет выводиться в виде чёрной рамки и распознанной буквы над ней.

Результаты. С помощью объявленных параметров обучаем модель на тренировочной выборке. На рис. 4 показаны результаты обучения модели. Процесс занял 28 минут на 10 подходах.

```
history = model.fit(train,label,epochs=10,validation_data=(test,test_label))

Epoch 1/10
/usr/local/lib/python3.10/dist-packages/keras/src/backend.py:5727: UserWarning: ``sparse_categorical_crossentropy`` received ``from_logits=True``,
output, from_logits = _get_logits(
858/858 [=====] - 171s 197ms/step - loss: 0.5237 - accuracy: 0.8366 - val_loss: 0.1583 - val_accuracy: 0.9428
Epoch 2/10
858/858 [=====] - 170s 198ms/step - loss: 0.0405 - accuracy: 0.9865 - val_loss: 0.1695 - val_accuracy: 0.9381
Epoch 3/10
858/858 [=====] - 168s 195ms/step - loss: 0.0291 - accuracy: 0.9897 - val_loss: 0.4449 - val_accuracy: 0.8581
Epoch 4/10
858/858 [=====] - 168s 196ms/step - loss: 0.0285 - accuracy: 0.9905 - val_loss: 0.5009 - val_accuracy: 0.8858
Epoch 5/10
858/858 [=====] - 171s 199ms/step - loss: 0.0189 - accuracy: 0.9944 - val_loss: 0.4315 - val_accuracy: 0.8839
Epoch 6/10
858/858 [=====] - 167s 195ms/step - loss: 0.0216 - accuracy: 0.9930 - val_loss: 0.6361 - val_accuracy: 0.8603
Epoch 7/10
858/858 [=====] - 169s 197ms/step - loss: 0.0197 - accuracy: 0.9937 - val_loss: 0.8175 - val_accuracy: 0.8253
Epoch 8/10
858/858 [=====] - 169s 197ms/step - loss: 0.0198 - accuracy: 0.9941 - val_loss: 0.8957 - val_accuracy: 0.7796
Epoch 9/10
858/858 [=====] - 169s 197ms/step - loss: 0.0195 - accuracy: 0.9941 - val_loss: 0.2757 - val_accuracy: 0.9407
Epoch 10/10
858/858 [=====] - 167s 195ms/step - loss: 0.0047 - accuracy: 0.9985 - val_loss: 0.4476 - val_accuracy: 0.9145
```

Рис. 4. Результаты обучения модели

Модель выдала высокий результат точности на тестовой выборке и составила 99%, что подтверждает высокую эффективность предложенной сверточной нейронной сети (CNN).

Для наглядного сравнения точности предложенного решения с другими методами распознавания жестов представлены в табл. 1.

Таблица 1

Сравнение методов распознавания жестов

Метод	Точность
Сверточная нейронная сеть (CNN)	99%
Классическая ИНС	89%
Логистическая регрессия	84%
CNN-LSTM	96,4%

CNN показала высокую точность (99%), что объясняется способностью сети эффективно выявлять пространственные признаки жестов и устойчивостью к шуму и вариациям освещения.

Классическая ИНС и логистическая регрессия продемонстрировали значительно более низкие показатели (89% и 84% соответственно), так как не учитывают пространственную структуру изображений [20].

CNN-LSTM достигла точности 96,4%, однако требует больше вычислительных ресурсов и сложнее в реализации [9]

Преимущества предложенного алгоритма:

- ♦ Высокая точность и надежность даже при сложных условиях освещения и частичной видимости пальцев.
- ♦ Реальное время обработки: интеграция с OpenCV и MediaPipe позволяет оперативно отображать результаты видеопотока.
- ♦ Сравнение с другими методами демонстрирует, что предложенная CNN превосходит классические ИНС и близка по результату к современным гибридным решениям при меньших вычислительных затратах.
- ♦ Устойчивость к вариациям жестов: система корректно распознает буквы при разных ракурсах, положениях рук и частичном перекрытии пальцев.

Предложенная модель готова к практическому применению в системах человеко-машинного взаимодействия, обеспечивая высокую точность распознавания русского жестового языка и минимальное запаздывание в обработке видеопотока. Дальнейшие работы будут направлены на увеличение объёма и разнообразия обучающих данных, интеграцию временных архитектур для анализа динамических жестов и повышение устойчивости к внешним помехам.

Выводы. Разработан и реализован алгоритм распознавания русского жестового языка (РЖЯ) на базе сверточной нейронной сети (CNN), интегрированной с системой отслеживания рук. Предложенное решение позволяет идентифицировать положение кисти и отдельных пальцев с высокой точностью, обеспечивая корректное распознавание жестов при частичном перекрытии пальцев, различных ракурсах и изменениях освещения.

Проведено обучение и тестирование модели на специализированном датасете, включающем 430 изображений, охватывающих буквы алфавита РЖЯ с различными ракурсами и условиями освещения. Точность распознавания на тестовой выборке составила 99 %, что подтверждает высокую эффективность архитектуры CNN в задачах пространственного анализа изображений жестов.

Практическая значимость исследования заключается в возможности использования предложенного алгоритма для:

- ♦ создания интерактивных систем коммуникации для глухонемых;
- ♦ разработки реабилитационных приложений и образовательных платформ;
- ♦ внедрения в интерфейсы виртуальной и дополненной реальности (VR/AR), где требуется распознавание жестов в реальном времени.

Научная новизна работы заключается в интеграции маркерного отслеживания рук с архитектурой CNN для распознавания жестов, что обеспечивает высокую точность и стабильность распознавания в условиях изменяющейся среды, а также в сравнительном анализе с классическими методами и гибридными CNN-LSTM моделями, подтверждающем преимущество предложенной модели.

Перспективы дальнейших исследований включают:

- ♦ расширение и усложнение датасета за счет включения динамических жестов и словосочетаний;
- ♦ интеграцию временных архитектур (LSTM, TCN) для распознавания динамических жестов и предложений;
- ♦ оптимизацию вычислительных процессов для работы на мобильных и встроженных устройствах;
- ♦ разработку алгоритмов автоматической коррекции ошибок распознавания и повышения устойчивости к шумам и нестандартным положениям руки.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. World Health Organization. Deafness and hearing loss. – 2023. – URL: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss> (дата обращения: 10.01.2026).
2. Дialectы русского жестового языка. – URL: <https://vuginfo.ru/vog/2015/05/30/rslidioms/> (дата обращения: 10.02.2024).
3. Cruz M., et al. Sign language recognition: A review // International Journal of Computer Vision. – 2020.
4. Kim H., et al. Marker-based hand gesture recognition systems: A review // Sensors. – 2022.
5. Foteinos K., et al. Visual Hand Gesture Recognition with Deep Learning: A Comprehensive Review of Methods, Datasets, Challenges and Future Research Directions // arXiv preprint arXiv:2507.04465. – 2025.
6. Zhou Y., et al. CNN-LSTM architecture for sign language recognition // IEEE Access. – 2021.
7. Wang L., et al. Spatio-temporal networks for dynamic hand gesture recognition // Neurocomputing. – 2024.
8. Garcia F., et al. Robust CNN models with data augmentation for hand gesture recognition // Pattern Recognition. – 2025.
9. Baihan A., Alutaibi A.I., Alshehri M., et al. Sign language recognition using modified deep learning network and hybrid optimization: a hybrid optimizer (HO) based optimized CNNSa-LSTM approach // Sci Rep. – 2024. – Vol. 14. – Article 26111. – DOI: 10.1038/s41598-024-76174-7.
10. Kumar S., et al. Vision-based markerless hand gesture recognition using deep learning // IEEE Transactions on Multimedia. – 2022.
11. Safa Ameur, Mohamed Ali Mahjoub, Anouar Ben Khalifa. Spatial-temporal generative network based on deep long short-term memory autoencoder for hand skeleton data sequences reconstruction and recognition // Engineering Applications of Artificial Intelligence. – 2025. – Vol. 161. – Part C. – Article 112289.

12. Al Mudawi N., Ansar H., Alazeb A., Aljuaid H., AlQahtani Y., Algarni A., Jalal A., Liu H. Innovative healthcare solutions: robust hand gesture recognition of daily life routines using 1D CNN // *Front. Bioeng. Biotechnol.* – 2024. – Vol. 12. – Article 1401803. – DOI: 10.3389/fbioe.2024.1401803.
13. Patel R., et al. Efficient CNN models for embedded hand gesture recognition // *Journal of Real-Time Image Processing.* – 2023.
14. Буйначев С.К., Боклаг Н.Ю. Основы программирования на языке Python. – Екатеринбург: Изд-во Урал. ун-та, 2014. – 91 с.
15. Жерон О. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow. – М.: Диалектика, 2018. – 688 с.
16. Маккини У. Python и анализ данных: пер. с англ. А.А. Слинкина. – М.: ДМК Пресс, 2020. – 540 с.
17. Сем В. Глубокое обучение: легкая разработка проектов на Python. – СПб.: Питер, 2021. – 352 с.
18. Collette A. H5Py Documentation. Release Sep 06, 2019. – 75 p.
19. Учебник «Машинное зрение». – Т. 2. – URL: <https://copter-space.gitbook.io/uchebnik-mashinnoe-zrenie-tom-2> (дата обращения: 10.04.2025).
20. Esan A., Okomba N., Adebisi T., et al. Comparing CNN and Logistic Regression Results for Automated Detection Tasks. – URL: <https://www.researchgate.net/publication/382880766> (дата обращения: 10.11.2025).

REFERENCES

1. World Health Organization. Deafness and hearing loss, 2023. Available at: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss> (accessed 10 January 2026).
2. Dialekty russkogo zhestovogo yazyka [Dialects of russian sign language]. Available at: <https://voginfo.ru/vog/2015/05/30/rsliidioms/> (accessed 10 February 2024).
3. Cruz M., et al. Sign language recognition: A review, *International Journal of Computer Vision*, 2020.
4. Kim H., et al. Marker-based hand gesture recognition systems: A review, *Sensors*, 2022.
5. Foteinos K., et al. Visual Hand Gesture Recognition with Deep Learning: A Comprehensive Review of Methods, Datasets, Challenges and Future Research Directions, *arXiv preprint arXiv:2507.04465*, 2025.
6. Zhou Y., et al. CNN-LSTM architecture for sign language recognition, *IEEE Access*, 2021.
7. Wang L., et al. Spatio-temporal networks for dynamic hand gesture recognition, *Neurocomputing*, 2024.
8. Garcia F., et al. Robust CNN models with data augmentation for hand gesture recognition, *Pattern Recognition*, 2025.
9. Baihan A., Alutaibi A.I., Alshehri M., et al. Sign language recognition using modified deep learning network and hybrid optimization: a hybrid optimizer (HO) based optimized CNNSa-LSTM approach, *Sci Rep.*, 2024, Vol. 14, Article 26111. DOI: 10.1038/s41598-024-76174-7.
10. Kumar S., et al. Vision-based markerless hand gesture recognition using deep learning, *IEEE Transactions on Multimedia*, 2022.
11. Sifa Ameur, Mohamed Ali Mahjoub, Anouar Ben Khalifa. Spatial-temporal generative network based on deep long short-term memory autoencoder for hand skeleton data sequences reconstruction and recognition, *Engineering Applications of Artificial Intelligence*, 2025, Vol. 161, Part C, Article 112289.
12. Al Mudawi N., Ansar H., Alazeb A., Aljuaid H., AlQahtani Y., Algarni A., Jalal A., Liu H. Innovative healthcare solutions: robust hand gesture recognition of daily life routines using 1D CNN, *Front. Bioeng. Biotechnol.*, 2024, Vol. 12, Article 1401803. DOI: 10.3389/fbioe.2024.1401803.
13. Patel R., et al. Efficient CNN models for embedded hand gesture recognition, *Journal of Real-Time Image Processing*, 2023.
14. Буynачев С.К., Боклаг Н.Ю. Osnovy programmirovaniya na yazyke Python [Fundamentals of programming in Python]. Ekaterinburg: Izd-vo Ural. un-ta, 2014, 91 p.
15. Zheron O. Prikladnoe mashinnoe obuchenie s pomoshch'yu Scikit-Learn i TensorFlow [Applied machine learning with Scikit-Learn and TensorFlow]. Moscow: Dialektika, 2018, 688 p.
16. Makini U. Python i analiz dannykh [Python and data analysis]: transl. from Engl. by A.A. Slinkina. Moscow: DMK Press, 2020, 540 p.
17. Set V. Glubokoe obuchenie: legkaya razrabotka proektov na Python [Deep learning: easy Python project development]. Saint Petersburg: Piter, 2021, 352 p.
18. Collette A. H5Py Documentation. Release Sep 06, 2019, 75 p.
19. Учебник «Машинное зрение» [Textbook "Machine Vision"]. Vol. 2. Available at: <https://copter-space.gitbook.io/uchebnik-mashinnoe-zrenie-tom-2> (accessed 10 April 2025).
20. Esan A., Okomba N., Adebisi T., et al. Comparing CNN and Logistic Regression Results for Automated Detection Tasks. Available at: <https://www.researchgate.net/publication/382880766> (accessed 10 November 2025).

Бондарева Валерия Евгеньевна – Южно-Российский государственный политехнический университет (НПИ) имени М.И. Платова; e-mail: bondareva.valeria2011@yandex.ru; г. Новочеркасск, Россия; кафедра «Информационные и измерительные системы и технологии»; бакалавр.

Черноморова Татьяна Семеновна – Южно-Российский государственный политехнический университет (НПИ) имени М.И. Платова; e-mail: tatyana.chernomorova@gmail.com; г. Новочеркасск, Россия; кафедра «Информационные и измерительные системы и технологии»; доцент; <https://orcid.org/0009-0003-2457-5263>.

Кривцун Анастасия Витальевна – Южно-Российский государственный политехнический университет (НПИ) имени М.И. Платова; e-mail: anastasia.srstu@gmail.com; г. Новочеркасск, Россия; кафедра «Информационные и измерительные системы и технологии»; доцент; <https://orcid.org/0000-0003-0332-4695>.

Абдулкарем Абейр – Южно-Российский государственный политехнический университет (НПИ) имени М.И. Платова; e-mail: Abeerabdulsalam15@gmail.com; г. Новочеркасск, Россия; кафедра «Информационные и измерительные системы и технологии»; аспирант; ORCID iD: 0009-0003-1892-9408.

Bondareva Valeria Evgenyevna – Platov South-Russian State Polytechnic University (NPI); e-mail: bondareva.valeria2011@yandex.ru; Novocherkassk, Russia; the Department “Information and Measurement Systems and Technologies”; bachelor.

Chernomorova Tatyana Semyonovna – Platov South-Russian State Polytechnic University (NPI); e-mail: tatyana.chernomorova@gmail.com; Novocherkassk, Russia; the Department “Information and Measurement Systems and Technologies”; associate professor; <https://orcid.org/0009-0003-2457-5263>.

Krivtsun Anastasia Vitalyevna – Platov South-Russian State Polytechnic University (NPI); e-mail: anastasia.srstu@gmail.com; Novocherkassk, Russia; the Department “Information and Measurement Systems and Technologies”; associate professor; <https://orcid.org/0000-0003-0332-4695>.

Abdulkarem Abeer – Platov South-Russian State Polytechnic University (NPI); e-mail: Abeerabdulsalam15@gmail.com; Novocherkassk, Russia; of the Department “Information and Measurement Systems and Technologies,” postgraduate student; ORCID iD: 0009-0003-1892-9408.

ПРАВИЛА ОФОРМЛЕНИЯ РУКОПИСЕЙ

С полными требованиями к рукописи, перечнем сопроводительных документов и образцом оформления статьи можно ознакомиться на сайте журнала:

https://izv-tn.tti.sfedu.ru/index.php/izv_tn/authorrequirements.

1. Объём статьи – от 12 до 20 страниц формата А4, включая рисунки, таблицы и список литературы (для обзорных статей – от 20 до 30 страниц). Основной текст статьи должен быть набран шрифтом Times New Roman размером 10 пунктов с одинарным интервалом.

2. Названию статьи предшествует индекс УДК, соответствующий заявленной теме.

3. Текст статьи начинается с названия статьи (на русском и английском языках), фамилии, имени и отчества автора (полностью) и снабжается аннотацией на русском и английском языках объёмом **200-250 слов**.

Аннотация должна содержать: актуальность исследования (постановка проблемы); цель и задачи работы; используемые методы и подходы; – основные результаты (с конкретикой, без общих формулировок); выводы и значимость работы.

4. Статья должна быть структурирована и включать: введение, постановку задачи, методику исследования, результаты и их обсуждение (или вычислительный эксперимент), выводы.

5. Все графические материалы (графики, диаграммы, фотографии) должны быть пронумерованы и снабжены подписями. Таблицы также нумеруются и имеют названия. Формулы должны быть набраны в редакторе формул.

6. Наличие библиографического списка на русском и английском языках обязательно. Все источники указываются в тексте в порядке возрастания. Должно быть не менее **20 источников** (для обзорных статей – не менее 40): 75% источников – не старше 10 лет; 25% источников – допускаются старше 10 лет (классические/базовые работы). Ссылки оформляются по ГОСТ Р 7.0.5-2008, в тексте заключаются в квадратные скобки.

7. Оригинальность текста статьи должна составлять не менее 70% по результатам проверки в системе антиплагиата. Объём самоцитирования не должен превышать 20% от общего объёма статьи. Автор обязан предоставить справку-отчёт системы антиплагиат (желательно **antiplagiat.ru**).

8. Статья сопровождается сведениями об авторе(ах) (фамилия, имя, отчество, учёное звание, должность, место работы, адрес, электронный адрес и номер телефона) на русском и английском языках, а также необходимыми сопроводительными документами в соответствии с требованиями журнала.

9. Срок рассмотрения статьи, соответствующей всем требованиям, составляет до 3 месяцев, с момента направления статьи на рецензирование.

10. Публикация в журнале для авторов бесплатна. Редакция не взимает плату с авторов за подготовку, размещение и печать материалов.

Статьи, не соответствующие требованиям журнала, возвращаются авторам на доработку до направления на рецензирование.