

ISSN 2223-2966



СОВРЕМЕННАЯ НАУКА:
АКТУАЛЬНЫЕ ПРОБЛЕМЫ
ТЕОРИИ И ПРАКТИКИ

ЕСТЕСТВЕННЫЕ И ТЕХНИЧЕСКИЕ НАУКИ

№ 5-2 2026 (МАЙ)

Учредитель журнала
Общество с ограниченной ответственностью
«НАУЧНЫЕ ТЕХНОЛОГИИ»

Журнал издается с 2011 года.

Редакция:

Главный редактор
А.В. Царегородцев
Выпускающий редактор
Ю.Б. Миндлин
Верстка
М.А. Комарова

Подписной индекс издания
в каталоге агентства «Пресса России» — 80016
В течение года можно произвести подписку
на журнал непосредственно в редакции.

Издатель:

Общество с ограниченной ответственностью
«Научные технологии»

Адрес редакции и издателя:
109443, Москва, Волгоградский пр-т, 116-1-10
Тел/факс: 8(495) 142-8681
E-mail: redaktor@nauteh.ru
<http://www.nauteh-journal.ru>

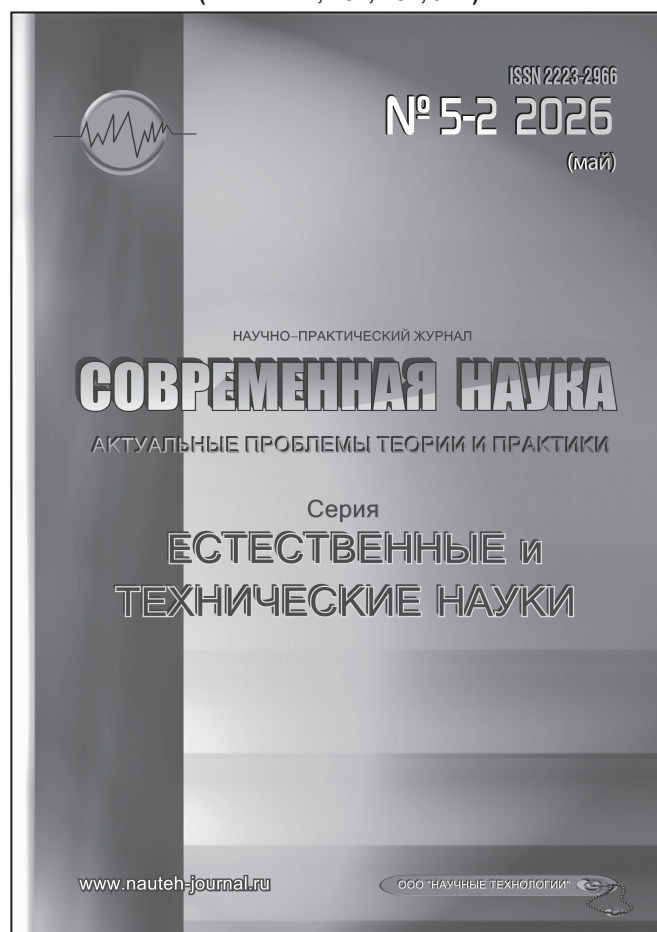
Журнал зарегистрирован Федеральной службой
по надзору в сфере массовых коммуникаций,
связи и охраны культурного наследия.

Свидетельство о регистрации
ПИ № ФС 77-44912 от 04.05.2011 г.

Научно-практический журнал

Scientific and practical journal

(BAK — 1.1.2, 1.5.x, 2.3.x, 3.1.x)



В НОМЕРЕ:

ОБЩАЯ БИОЛОГИЯ, ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ,
КЛИНИЧЕСКАЯ МЕДИЦИНА

Авторы статей несут полную ответственность
за точность приведенных сведений, данных и дат.

При перепечатке ссылка на журнал
«Современная наука:
Актуальные проблемы теории и практики» обязательна.

Журнал отпечатан в типографии
ООО «КОПИ-ПРИНТ» тел./факс: (495) 973-8296
Подписано в печать 29.05.2026 г. Формат 84x108 1/16
Печать цифровая Заказ № 0000 Тираж 2000 экз.

ISSN 2223-2966



Редакционный совет

Атаев Алевдин Рашитханович — д.м.н., профессор, Дагестанский государственный медицинский университет

Безруких Марьям Моисеевна — д.б.н., профессор, Институт возрастной физиологии РАО

Бекетов Сергей Валериевич — д.б.н., ФГБНУ НИИ Пушного звероводства и кролиководства имени В.А. Афанасьева

Белых Владимир Иванович — Д.м.н., доцент, Алтайский государственный медицинский университет Министерства здравоохранения Российской Федерации

Грачев Николай Николаевич — доктор высшей степени в области технических наук (DoctorHabilitatus), профессор, Московский государственный институт электроники и математики НИУ ВШЭ (технический университет)

Гусева Анна Ивановна — д.т.н., профессор, Национальный исследовательский ядерный университет «МИФИ»

Зубкова Валентина Михайловна — д.б.н., профессор, Российский государственный социальный университет

Каллаев Нажмудин Омаркадиевич — д.м.н., профессор, Дагестанский государственный медицинский университет

Квасов Андрей Иванович — д.т.н., профессор, Восточно-Казахстанский государственный технический университет им. Д. Серикбаева

Корнеев Андрей Матиславович — д.т.н., профессор, Липецкий государственный технический университет

Корягина Наталья Александровна — д.м.н., доцент, Пермский государственный медицинский университет им. ак. Е.А.Вагнера Министерства здравоохранения Российской Федерации

Кравец Бронислава Борисовна — д.м.н., профессор, Воронежский государственный медицинский университет им. Н.Н. Бурденко Министерства здравоохранения Российской Федерации

Кулик Сергей Дмитриевич — д.т.н., доцент, Национальный исследовательский ядерный университет «МИФИ»

Матвеев Всеволод Борисович — д.м.н., профессор, ФГБУ РОНЦ им. Н.Н. Блохина, член-корреспондент РАН

Миндлин Юрий Борисович — к.э.н., доцент, Московская государственная академия ветеринарной медицины и биотехнологии им. К.И. Скрябина

Надежкин Сергей Михайлович — д.б.н., профессор, Всероссийский НИИ селекции и семеноводства овощных культур Россельхозакадемии

Овезов Алексей Мурадович — д.м.н., доцент, ГБУЗ МО Московский областной научно-исследовательский клинический институт им. М.Ф. Владимирского

Олейникова Светлана Александровна — д.т.н., доцент, Воронежский государственный технический университет

Рахимов Ильгизар Ильясович — д.б.н., профессор, Казанский (Приволжский) федеральный университет

Ромашкова Оксана Николаевна — д.т.н., профессор, Московский городской педагогический университет

Симаков Юрий Георгиевич — д.б.н., профессор, Московский государственный университет им. К.Г. Разумовского (ПКУ)

Симоненков Алексей Павлович — д.м.н., профессор, независимый эксперт

Трапезов Олег Васильевич — д.б.н., в.н.с., ФИЦ «Институт цитологии и генетики СО РАН»

Федорова Оксана Ивановна — д.б.н., доцент, Московская государственная академия ветеринарной медицины и биотехнологии им. К.И. Скрябина

Харитонов Михаил Анатольевич — д.м.н., профессор, Военно-медицинская академия имени С.М. Кирова

Царегородцев Анатолий Валерьевич — д.т.н., профессор, Московский государственный лингвистический университет redaktor3@nauteh.ru

СОДЕРЖАНИЕ

CONTENTS

Общая биология

Герасименко А.А., Горох А.М., Писанов Р.В., Водопьянов А.С. — Классификация и автоматическое типирование штаммов вируса бешенства на территории РФ на основе различий в строении n гена
Gerasimenko A., Gorokh A., Pisanov R., Vodopyanov A. — Automatic typing of Lyssavirus rabies strains in Russian Federation by n gene building differences7

Калинникова Т.Б., Егорова А.В., Гатиятуллина А.Ф., Кузнецова Т.В., Князев И.В., Петров А.М. — Исследование токсикологических характеристик термически обработанного осадка сточных вод
Kalinnikova T., Egorova A., Gatiyatullina A., Kuznetsova T., Knyazev I., Petrov A. — Study of the toxicological characteristics of the thermomechanically treated sewage sludge13

Маркин Е.М., Терехова М.В., Отрубянников Е.В. — Твердые дисперсные материалы как экологически устойчивая форма аккумуляции теплоты
Markin E., Terekhova M., Otrubyanikov E. — Granular solid materials as an environmentally sustainable form of heat storage17

Трунова С.А., Нурмагомедова С.Г., Магомедова У.А., Даудова Р.Д., Сулейманова Р.Г. — Концепция «Единое здоровье» (One Health) в системе экологической безопасности
Trunova S., Nurmagomedova S., Magomedova U., Daudova R., Suleymanova R. — The one health concept in the environmental safety system25

Информатика, вычислительная техника и управление

Адаев Н.О., Пожидаев Н.А., Маркарян П.Г., Таранец М.Е., Малахов С.В. — Методика функциональной эмуляции протокола New IP в среде GNS3 и экспериментальная оценка эффективности инкапсуляции для IoT-сегментов перспективных сетей

Adaev N., Pozhidaev N., Markaryan P., Taranets M., Malakhov S. — A methodology for functional emulation of the New IP protocol in the GNS3 environment and an experimental evaluation of encapsulation efficiency for IoT segments of promising networks29

Алькаттан Самех — Структурные модели LLM-агентов и их ограничения с точки зрения информационной безопасности: обзор и сравнительный анализ
Alkattan Sameh — Structural models of LLM agents and their security limitations: a review and comparative analysis35

Апана Кеннет Айнбуно, Шичкина Ю.А. — Эффективная обработка малых данных в трёхмерной объёмной сегментации нейроизображений: ограничение параметров как структурный регуляризатор
Apana Kenneth Ayinbuno, Shichkina Yu. — Efficient small data processing in three-dimensional volumetric segmentation of neuroimages: parameter constraint as a structural regulator42

Васин Н.А., Шмаков В.Э. — Система автоматизированного взаимодействия с заказчиком в управлении проектами
Vasin N., Shmakov V. — System of automated interaction with customer in project management48

Воробьёв Е.А., Пикина Н.Е., Работает Г.А. — Особенности дообучения энкодерных больших языковых моделей для работы с новой предметной областью используя функцию потерь Triplet Loss
Vorobyov E., Pikina N., Rabotaev G. — Features of fine-tuning encoder-based large language models for a new subject domain using the Triplet Loss function53

Воронин С.А. — Проблемы и перспективы безопасного использования персональных данных в системах искусственного интеллекта: комплексный анализ правовых, технических и методологических аспектов

Voronin S. — Problems and prospects of the safe use of personal data in artificial intelligence systems: a comprehensive analysis of legal, technical and methodological aspects60

Глухов А.А. — Об определении требований по защите информации в доверенных программно-аппаратных комплексах с учетом результативности мониторинга информационной безопасности
Glukhov A. — On the definition of information security requirements in trusted software and hardware complexes, taking into account the effectiveness of information security monitoring.....69

Горячкин Б.С., Панов Г.Д. — Эргономическая оценка авто-генерации UI с использованием AI-агента и MCP-сервера
Goryachkin B., Panov G. — Ergonomic evaluation of UI auto-generation using an AI agent and an MCP server.....77

Горячкин Б.С., Сыса П.В. — Анализ времени отклика на поисковые запросы в системах поиска научных публикаций
Goryachkin B., Sysa P. — Analysis of query response time in scientific publication retrieval systems83

Гриценко Д.А., Парфентьев К.В. — Обзор методов искусственного интеллекта для поиска аномалий в распределенных системах
Gritsenko D., Parfentiev K. — Review of artificial intelligence methods for detecting anomalies in distributed computers.....92

Евдошенко О.И. — Исследование архитектуры системы экологического мониторинга Арктики
Evdoshenko O. — Research of the architecture environmental monitoring system of the Arctic.97

Есипов И.В., Староверов И.Н. — Многолетний анализ и предиктивное моделирование климатических изменений в Московском регионе с использованием методов машинного обучения
Esipov I., Staroverov I. — Long-term analysis and predictive modeling of climate change in the Moscow region using machine learning methods 100

Катасонов Д.Н., Москалева В.М. — Алгоритм автоматического поддержания расстояния до обрабатываемой подложки в системах лазерной обработки

Katasonov D., Moskaleva V. — Algorithm for automatically maintaining the distance to the processed substrate in laser processing systems .. 111

Клементьева О.В., Чмилевская О.С. — Факторы, снижающие результативность управления устареванием оборудования в организациях
Klementyeva O., Chmylevskaya O. — Factors reducing the effectiveness of equipment obsolescence management in organizations 116

Козлов А.В., Кочемиров М.А., Ушаков Д.П. — Количественная оценка влияния понижения разрешения растровых изображений на этапе декодирования на потребление памяти, поведение кеша и плавность интерфейса во Flutter
Kozlov A., Kochemirov M., Ushakov D. — Quantative assessment of the impact of decode-time downscaling of raster images on memory consumption, image cache behaviour and user interface smoothness in Flutter 129

Конде Абдул Карим — Исследование декларативных подходов к выбору представлений для повышения эффективности аналитической обработки данных
Conde Abdoul Karim — The study of declarative approaches to the choice of views for improving the efficiency of analytical data processing 134

Ласькова Ю.С. — Принятие решений по техническому обслуживанию с использованием интеллектуальной прогностики в рамках единой информационной системы принятия решений на авиационном предприятии
Laskova Yu. — Maintenance decision-making using intelligent forecasting within the framework of a unified decision-making information system at an aviation enterprise 139

Малеев О.Г. — Гибридный подход к построению нейросетевой диалоговой системы с использованием графа переходов для улучшения обработки эллипсисов
Maleev O. — A hybrid approach to constructing a neural network dialogue system using a transition graph to improve ellipsis handling 145

Мирсагатов Д.Ф. — Сравнительный анализ нейросетевых моделей суперразрешения изображений

Mirsagatov D. — Comparative analysis of neural network models of image super resolution 148

Молькова Л.Ю. — Модель устойчивого маркирования сетевых потоков для повышения защищённости частично контролируемых сетей
Molkova L. — A model for robust network flow marking for enhancing the security of partially controlled networks 153

Нгула Бринч-Жофрей — Глубокое обучение с подкреплением: алгоритмы, приложения и вызовы
Ngoula Brintch-Geoffrey — Deep reinforcement learning: algorithms, applications, and challenges 161

Петров Д.С., Петров М.С. — Нейронная сеть для планирования перемещения автономного робота
Petrov D., Petrov M. — Neural network for path planning of an autonomous robot 165

Пожар К.В. — Принцип человеко-машинного консенсуса при интеллектуальной поддержке принятия решений в автоматизированных системах нелинейного управления доставкой инсулина
Pozhar K. — The principle of human-machine consensus in intelligent decision support in automated insulin delivery non-linear control systems 170

Поташин А.А. — Помехоустойчивость OFDM-модуляции при передаче видеосигнала в условиях ретранслированных помех: имитационное исследование в GNU Radio
Potashin A. — Noise-resistant OFDM-modulation of video signal transmission under conditions of retranslated noise: simulation study in GNU Radio 175

Тоичкин Н.А. — Разработка информационной системы видеоаналитики для обнаружения дефектов конвейерной ленты транспортировки руды
Toichkin N. — Development of a video analytics information system for detecting defects in ore conveyor belts 183

Тшипунгу Эдуарду Афонсу, Шатунова В.М. — Интегральный индекс загрязнения по спутниковым данным: автоматизация расчёта

и визуализации для расследования аварии на хвостохранилище Катока

Tchipungu Eduardo Afonso, Shatunova V. — Integrated pollution index via satellite data: automating calculation and visualization for the investigation of the Catoca tailings dam accident 189

Узбеков Р.Ш., Челноков В.В. — Математическое моделирование и оптимизация энергоэффективных режимов криогенного разделения воздуха в условиях технологической неопределённости
Uzbekov R., Chelnokov V. — Mathematical modeling and optimization of energy-efficient cryogenic air separation modes under conditions of technological uncertainty 197

Цай Цин — Разработка системы симуляции работы алгоритмов управления электроэнергией в мультиагентных системах в рамках микрогрида на базе модели двойного аукциона
Tsai Tsin — Development of a simulation system for energy management algorithms in multi-agent systems within a microgrid based on the double auction model 203

Черенкова И.А., Кишкинова О.А., Кутликова И.В. — Обеспечение кибербезопасности на промышленных предприятиях с помощью искусственного интеллекта
Cherenkova I., Kishkinova O., Kutlikova I. — Ensuring cybersecurity in industrial enterprises using artificial intelligence 208

Черепенин В.А., Галушкина Д.В., Воробьев С.П. — Анализ архитектур и технологий Интернета вещей в распределённых вычислительных системах
Cherepenin V., Galushkina D., Vorobyov S. — Analysis of architectures and technologies of the Internet of Things in distributed computing systems 213

Шибиченко М.И. — Моделирование динамики потребительского спроса на основе интеграции эндогенных и экзогенных факторов с учетом их лагового воздействия
Shibichenko M. — Modelling the dynamics of consumer demand based on the integration of endogenous and exogenous factors, taking into account their lag effects 219

Клиническая медицина

Ерокина Н.Л., Миронов А.Ю., Захарова Н.Б., Парфенов А.К., Джагарян П.Д., Бахтеева Г.Р., Савельева С.С., Балакина Д.Д. — Диагностическая информативность биомаркеров при оценке воспалительных и репаративных процессов полости рта

Erokina N., Mironov A., Zakharova N., Parfenov A., Dzhagaryan P., Bakhteeva G.R., Savelyeva S., Balakina D. — Diagnostic informativeness of biomarkers in assessing inflammatory and reparative processes in the oral cavity 224

Котельников В.В., Целкович Л.С., Балтер Р.Б., Иванова Т.В., Ильченко О.А. — Современное состояние проблемы лечения эндометриом: противоречия, риски и перспективы персонализированной терапии

Kotelnikov V., Tselkovich L., Balter R., Ivanova T., Ilchenko O. — The current state of the endometrioma treatment problem: contradictions, risks and prospects for personalized therapy. 227

Кучеренко С.В., Корягина Н.А., Корягин В.С. — Клинико-инструментальные предикторы рецидива фибрилляции предсердий после катетерной абляции у пациентов, перенесших COVID-19: результаты проспективного сравнительного исследования

Kucherenko S., Koriagina N., Koriagin V. — Clinical and instrumental predictors of atrial fibrillation recurrence

after catheter ablation in COVID-19 survivors: a prospective comparative study. 232

Матвеев Д.В., Кузнецов М.Р., Дolidзе Д.Д., Лебедев С.С., Чебан С.В. — Влияние онкоассоциированной венозной патологии на качество жизни хронических онкологических больных и роль ранней терапии: проспективное исследование

Matveev D., Kuznetsov M., Dolidze D., Lebedev S., Cheban S. — Impact of cancer-associated venous pathology on quality of life in chronic oncological patients and the role of early therapy: a prospective study. 237

Сизов А.В. — Особенности ведения пациенток с гестационным сахарным диабетом и сопутствующим ожирением

Sizov A. — Features of management of patients with gestational diabetes mellitus and concomitant obesity. 245

Симанов И.В. — Зависимость частоты преждевременных и оперативных родов от наличия экстрагенитальной патологии у пациенток с преэклампсией различной степени тяжести

Simanov I. — Dependence of the frequency of premature and operative labor on the presence of extragenital pathology in patients with preeclampsia of various degrees of tension 250

Наши авторы 254

КЛАССИФИКАЦИЯ И АВТОМАТИЧЕСКОЕ ТИПИРОВАНИЕ ШТАММОВ ВИРУСА БЕШЕНСТВА НА ТЕРРИТОРИИ РФ НА ОСНОВЕ РАЗЛИЧИЙ В СТРОЕНИИ N ГЕНА

AUTOMATIC TYPING OF LYSSAVIRUS RABIES STRAINS IN RUSSIAN FEDERATION BY N GENE BUILDING DIFFERENCES

A. Gerasimenko
A. Gorokh
R. Pisanov
A. Vodopyanov

Summary. As a result of the study of 335 rabies virus genomes from the NCBI Virus and VGARus databases and sequenced on the basis of the «Rostov-on-Don Antiplague Scientific Research Institute» of Rospotrebnadzor, a unified classification of genomes by N gene's mutation profiles was developed, with the calculation of the optimal SNP threshold. 15 genetic clusters with unique marker mutations in N gene were identified. An author's Python program has been developed to validate the quality of nucleotide sequences of N gene and G gene to addition, automatically determine the genetic group, and obtain visual results for the virus genomes under study.

Keywords: rabies, Lyssavirus rabies, sequencing, typing, classification.

Герасименко Артем Александрович

младший научный сотрудник, ФКУЗ Ростовский-на-Дону
противочумный институт Роспотребнадзора
gerasimenko_aa@antiplague.ru

Горох Алевтина Михайловна

младший научный сотрудник, ФКУЗ Ростовский-на-Дону
противочумный институт Роспотребнадзора

Писанов Руслан Вячеславович

кандидат биологических наук, ведущий научный
сотрудник, ФКУЗ Ростовский-на-Дону противочумный
институт Роспотребнадзора

Водопьянов Алексей Сергеевич

кандидат медицинских наук, ведущий научный
сотрудник, ФКУЗ Ростовский-на-Дону противочумный
институт Роспотребнадзора

Аннотация. В результате исследования 335 геномов вируса бешенства из баз данных NCBI Virus, VGARus и секвенированных на базе ФКУЗ Ростовский-на-Дону противочумный институт Роспотребнадзора была разработана унифицированная классификация геномов по мутационным профилям в N гене, с вычислением оптимального порога разделения по SNP. Было выделено 15 генетических кластеров с уникальными маркерными мутациями в N гене у их представителей. Разработана авторская компьютерная программа на языке Python для валидации качества нуклеотидных последовательностей N и дополнительно G гена, автоматического определения генетической группы и получения наглядного результата по исследуемым геномам вируса.

Ключевые слова: бешенство, Lyssavirus rabies, секвенирование, типирование, классификация.

Бешенство — это заболевание, вызываемое нейротропным вирусом, которое регистрируется более чем в 150 странах и территориях по всему миру, за исключением Австралии и Антарктиды. Постоянный широкомасштабный мониторинг этой болезни является неотъемлемой частью комплекса противоэпизоотических и противоэпидемических мероприятий. На территории России почти во всех регионах регистрируются случаи бешенства, что делает актуальным совершенствование системы раннего выявления случаев заболевания и организации систематического лабораторного мониторинга циркулирующих штаммов вируса с использованием молекулярно-генетических методов.

Ген нуклеопротеина (N ген) вируса бешенства традиционно используется в качестве мишени для молекуляр-

ного типирования, так как он демонстрирует высокий уровень межштаммовой изменчивости при внутривидовой консервативности. Сравнение различных геновариантов может служить основой для понимания эволюции вируса, что представляет ценную информацию для разработки эффективной стратегии борьбы с ним. Это наглядно продемонстрировано на примере анализа штаммов, выделенных в Центральной Африке [1,2]. Молекулярно-генетические методы позволили установить факт трансграничного заноса инфекции в Забайкальский край из Монголии и Китая [3]. Исследование изолятов из центральной части России показало, что они образуют единый территориальный кластер, включающий Тверскую, Московскую и Тульскую области [4]. На основе структуры N гена возможно устанавливать филогенетические связи между изолятами возбудителя из различных источников.

В настоящее время существует несколько классификаций вируса бешенства, которые основаны на вариабельности его генома. Например, российские учёные предложили систему, согласно которой все варианты вируса делятся на шесть групп. Они провели филогенетический анализ исследованных вариантов, основываясь на структуре N гена [5,6]. Международная группа исследователей [7] (6 институтов из Франции, Африки, Австралии, США), в свою очередь, провела глубокий анализ 321 генома вируса бешенства, используя аналогичные методы. В результате они распределили изоляты, выделенные в разных регионах мира, на 60 различных кластеров.

За последнее время несколькими учеными были предложены различные программы для анализа и определения групповой принадлежности штаммов вируса бешенства. Nextstrain Rabies — платформа, предназначенная для филогенетического анализа геномных последовательностей в режиме реального времени. Она позволяет строить интерактивные деревья, которые отражают географическое распространение и ареал обитания возбудителя, основываясь на классификации Troupin et al., 2016. Последовательности автоматически загружаются из базы данных NCBI, а также могут вноситься пользователями вручную [8]. Предложен сервис RABV-GLUE, представляющий собой платформу с курируемой базой данных, которая позволяет по загружаемой последовательности определять кладу исследуемой пробы, идентифицировать завозные случаи и их источники, зоны длительной циркуляции вируса и пути его распространения (включая трансграничные) [9]. Апробирован сервис MADD OG — инструмент командной строки для автоматического определения генетических линий, реализующий принцип динамической филогенетической номенклатуры [10], который позволяет уточнить или выявить новые клады в уже существующих классификациях. MADD OG обеспечивает алгоритмическое разбиение клад на более мелкие филогенетические линии на основе эволюционных деревьев, задавая каждому кластеру уникальный код (иерархические обозначения типа A, A1, A1.1 и т.д.) по мере обнаружения новых ответвлений [9].

Большинство таких платформ представляют собой онлайн-сервисы, которые работают на зарубежных серверах и требуют специальной подготовки в области биоинформатики. При определении принадлежности штамма возбудителя к тому или иному кластеру существующие классификации предполагают построение филогенетических деревьев. При этом используются различные подходы и модели, что затрудняет воспроизведение результатов и изучение вновь выделяемых вирусов другими авторами. В отечественной практике на сегодняшний момент отсутствует унифицированный алгоритм и программа для автоматической и воспроизводимой онлайн классификации штаммов на кластеры

с учетом ключевых мутаций, имеющих эпидемиологическую или географическую значимость.

Разработка алгоритма, способного определять кластерную принадлежность вновь выделенных штаммов по уникальной структуре N гена, позволит существенно сократить время интерпретации результатов секвенирования и повысить воспроизводимость данных при проведении мониторинга. Особенно это актуально в условиях вспышек, требующих оперативного определения источника заражения и оценки путей распространения вируса [11]. Разработка отечественного программного решения, объединяющего выравнивание, выявление маркерных позиций и классификацию, является актуальной задачей на стыке биоинформатики, вирусологии и прикладной эпидемиологии.

Целью данной работы явилось изучение генетического разнообразия и разработка классификации с филогеографической характеристикой изолятов вируса бешенства, выделяемых на территории Российской Федерации.

Материалы и методы

В работе использованы геномы 335 штаммов вируса бешенства, выделенных с 1974 по 2024 год на 38 административных территориях Российской Федерации. Последовательности вирусов с известным годом, местом выделения и 100% покрытием N гена были получены из базы данных NCBI Virus (167), VGARus (160), а также использованы 8 геномов, секвенированных ранее на базе ФКУЗ Ростовский-на-Дону противочумный институт Роспотребнадзора. Анализируемые последовательности загружены из баз данных NCBI Virus с помощью NCBI Batch Entrez, а из VGARus — с помощью авторского скрипта. Последовательности N гена извлечены с помощью авторской программы Fragment Extractor, написанной на языке Java. Также для сравнения изучаемых геномов по международной классификации была сформирована база данных из геномов с известной принадлежностью к установленной генетической группе, упомянутых в статье Troupin et al., 2016 [7]. Выравнивание последовательностей выполнялось пакетом mafft на референсную последовательность N гена штамма вируса бешенства (NCBI Accession Number NC_001542) и с помощью авторского скрипта, разработанного на языке Python, были найдены нуклеотидные мутации.

Для изучаемых российских геномов авторским скриптом проводился кластерный анализ с визуализацией обнаруженных генетических групп. Порог по однонуклеотидным полиморфизмам (SNP) и число валидных кластеров подбирались с помощью elbow method и метрики silhouette score.

На языке Python с графической библиотекой `dearpygui` создана программа `RabiesAnalyzerRus`, позволяющая определять по загруженной последовательности принадлежность пробы к тем или иным кластерам с расшифровкой по предполагаемому географическому происхождению по разработанной в данном исследовании классификации и установленной международной. Алгоритм работы программы состоит в нахождении нуклеотидных мутаций и сравнении исследуемой последовательности с последовательностями N гена эталонных штаммов из баз данных, составленных из изученных штаммов и взятых из статьи по международной классификации Troupin et al., 2016 [7]. По таблицам, содержащим колонки «Проба — Мутации — Кластер», находится наиболее похожий «референсный» штамм, имеющий наибольшее количество совпадений по определенным мутациям с исследуемой пробой. Кластер и регион, к которому принадлежит такой штамм, признают предполагаемыми кластером и регионом исследуемого изолята.

Результаты и обсуждение

На первом этапе была построена матрица попарных SNP-различий между штаммами. Был реализован алгоритм с пороговой кластеризацией на основе минимальных остовных деревьев (MST) с биологически обоснованной фильтрацией для выделения генетически однородных групп штаммов по данным о мутациях. Для повышения биологической достоверности разделения на кластеры вначале также была применена двухэтапная фильтрация данных: во-первых, отбирались только кластеры размером не менее трех штаммов; во-вторых, из них выбирались те, которые содержали мутации, встречающиеся минимум у трех штаммов внутри кластера и которые полностью отсутствовали во всех других крупных кластерах.

Далее подбирался оптимальный порог SNP. Просматривалось множество возможных значений (все уникальные ненулевые расстояния по SNP между парами штаммов) и для каждого оценивалось качество получаемой кластерной структуры. Затем был применен *elbow method*: при каком пороге происходит резкое замедление роста числа валидных кластеров. Это значение интерпретировалось как оптимальный баланс: достаточно строгое разделение, чтобы отделить четкие группы, но не настолько, чтобы искусственно разбивать популяцию на мелкие группы. Таким образом, выбор порога — результат оптимизации, основанный на структуре данных, а не на субъективном предположении, позволяющий объективно определить количество генетически обособленных групп. Однако выбор только по точке излома не может быть достаточно точным, особенно если несколько порогов дают схожее число кластеров. Поэтому *elbow method* дополнился анализом *silhouette score* — метрики, оценивающей качество кластеризации по степени компактности компонентов внутри кластеров и разнесенности между ними. Из всех порогов, близких к точке локтя (*elbow*), выбирался тот, который максимизирует *silhouette score*. В итоге получен оптимальный порог, равный 6 SNP, необходимый для разделения штаммов на генетические кластеры (рис. 1).

На заключительном этапе строилось объединенное минимальное остовное дерево по всем валидным кластерам, а также формировалась полная разметка членства всех штаммов. Представители были разбиты на 15 групп, где каждая была названа по региону, изоляты которого встречались чаще всего. С помощью инструментов языка Python были найдены уникальные маркерные мутации в N гене для каждого установленного кластера, по которым представители отличаются друг от друга (табл. 1).

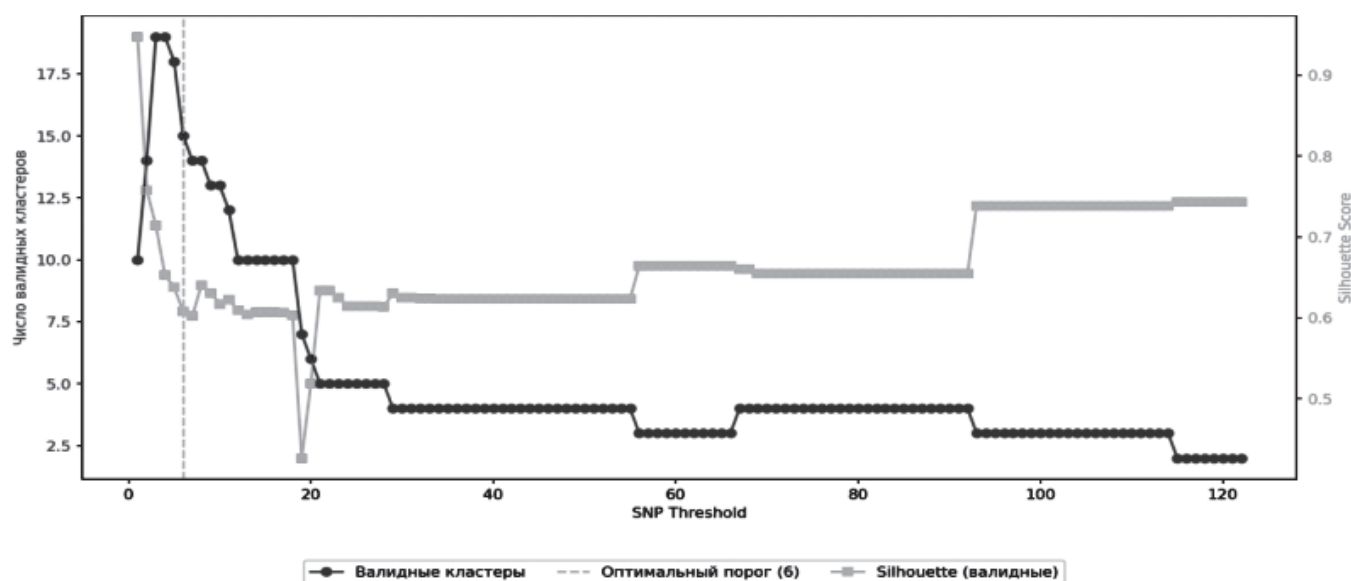


Рис. 1. Метод идентификации и валидации генетических кластеров по мутационным профилям N генов

Таблица 1.

Кластеры штаммов вируса бешенства, выявленные на основе полиморфизма N-гена

Кластер	Преимущественный регион распространения	Уникальные мутации
N1	Чукотка	A132T, T759C
N2	Бурятия	A387C, C1197T
N3	Бурятия / Забайкальский край	A165G
N4	Тыва	A760C, T1131G, A1328C
N5	Красноярский край	G924A
N6	Красноярский край / Хакасия	T198C, A456G, G804A, A1100G, T1338C
N7	Хакасия / Красноярский край	T570C
N8	Красноярский край / Алтай / Омская область	C1264A
N9	Омская область	C270T
N10	Омская / Новосибирская область	C621T, T954C, C1029A
N11	Центральная Россия	C375T
N12	Владимирская область	A906T, T1116C
N13	Черноземье	T942C
N14	Волгоградская область	T594C
N15	Крым	C45T, C270A

Таблица 1 демонстрирует распределение уникальных мутационных профилей 15 кластеров (от N1 до N15). Каждый кластер характеризуется определенным набором SNP (от 1 до 5). Стоит отметить наличие мутации C270A в кластере N15 и C270T в кластере N9, характерные для вирусов, выделенных на отдаленных территориях, таких как Республика Крым и Омская область. Возможно, данный полиморфизм играет определенную роль в адаптации вируса к тому или иному резервуару или ареалу. Примечательно также, что в одном макрорегионе (Бурятия, Красноярский край, Хакасия, Омская область), наблюдается циркуляция различных генетических групп вируса. Это может обуславливаться обширностью ареалов диких животных, разнообразием ландшафта, наличием большого числа обитающих животных — потенциальных резервуаров возбудителя — а также, возможно, присутствием трансграничного переноса штаммов, геноварианты которых характерны для соседних территорий.

В результате проведенного исследования получена биологически интерпретируемая структура популяции вирусов, поддерживающая гипотезу о существовании эволюционно обособленных линий. Данные о принад-

лежности изолятов к каждому кластеру были внесены в таблицу, включающую в себя колонки «Изолят-Регион-Мутации-Кластер», и была сформирована референсная база данных.

Был разработан подход для определения генотипа штамма без построения дендрограммы. Была создана программа RabiesAnalyzerRus, позволяющая по загружаемым последовательностям автоматически определять принадлежность проб к тому или иному кластеру с указанием региона преимущественного распространения определенного геноварианта. Инструмент включает в себя выравнивание на референсную последовательность единичных или нескольких геномов; определение мутаций в N гене каждого штамма; валидацию изучаемых последовательностей N и G гена (поиск стоп-кодонов, ошибок в каждом гене); сравнение профилей мутаций в N гене с заранее определенными у референсных штаммов из сформированных баз данных; определение генетических линий изолятов авторской методикой плюс соотнесение групп с международно установленными линиями (рис. 2).

В дальнейшем также планируется реализация классификации анализируемых последовательностей, исходя из строения не только N, но и G гена, поэтому в программу заранее была включена валидация обоих генов.

По окончании работы в указанной пользователем папке или директории по умолчанию будет находиться файл AboutProbes-{Дата исследования} формата tsv или json, где можно будет посмотреть список исследованных последовательностей с обнаруженными в них нуклеотидных мутаций в N гене, результат валидации по двум генам, название кластеров по авторской и международной классификации, к которому принадлежит определенный изолят и возможный регион, откуда он происходит. Те изоляты, которые в базе данных не образовали валидные кластеры планируется в дальнейшем использовать для дополнения, уточнения классификации и формирования генетических сублиний. Например, при поступлении в анализ последовательностей, которые будут наиболее близки к референсным, и полученная группа образует три штамма, формирующие валидный кластер с уникальными мутациями, будет сформирован валидный кластер с соответствующим наименованием по региону.

В ходе анализа также выяснено, что установленная по международной классификация линия Cosmopolitan-CA1 разбивается на 13 более мелких, штаммы которых входят в определенный регион страны и генетическую линию, где они отличаются друг от друга по набору ключевых мутаций. Штаммы линии Cosmopolitan-CA3 выделяются в Центральной России, а штаммы линии Arctic-related-A — в Чукотском автономном округе, далеком арктическом регионе нашей страны.

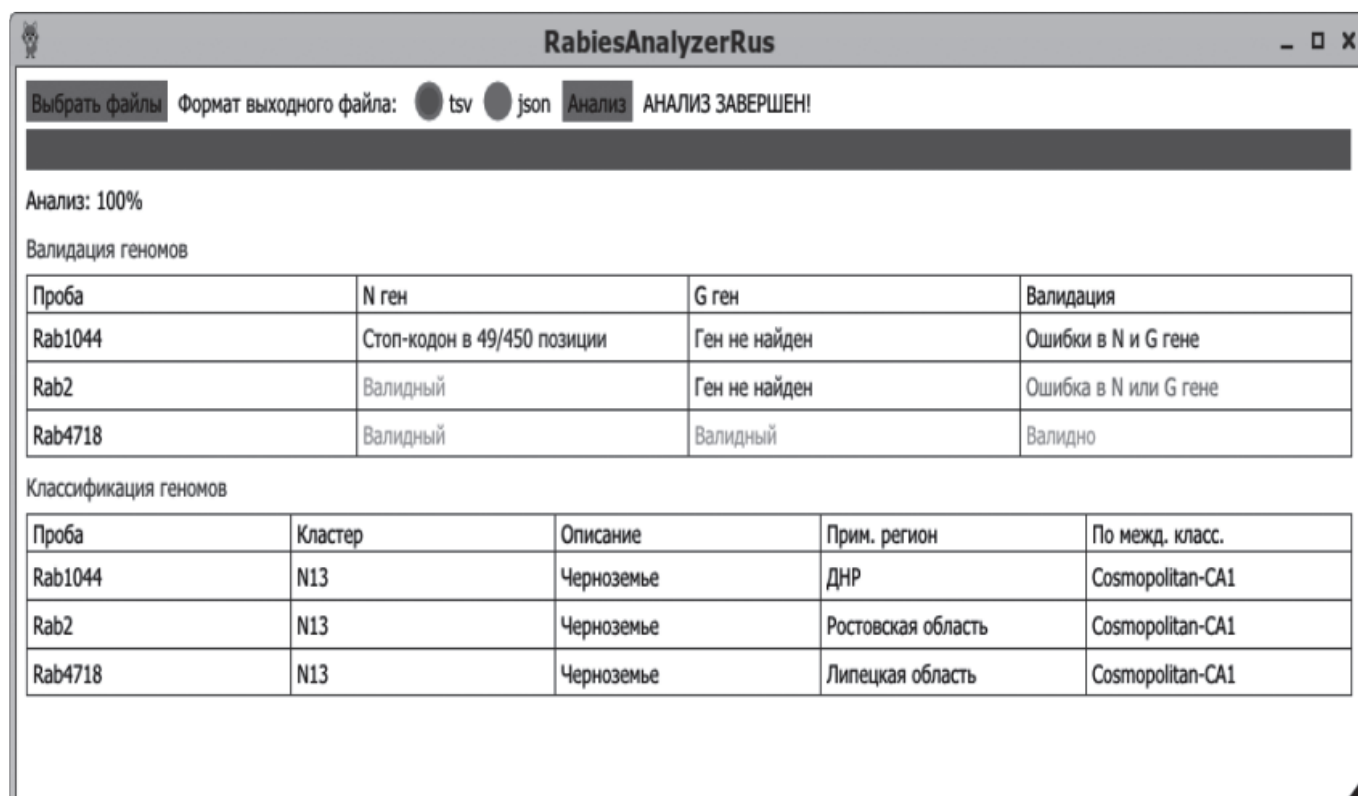


Рис. 2. Интерфейс работы программы при анализе геномов вируса бешенства

Заключение

Разработанная классификация расширяет представление о циркулировании уникальных геновариантов штаммов вируса бешенства, позволяет выявить линии, характерные как для макрорегионов Российской Федерации, так и для отдельных субъектов страны.

В результате проведенной работы предложен алгоритм классификации изучаемых штаммов возбудителя бешенства на кластеры с привязкой к географическо-

му происхождению; найдены маркерные мутации, характерные для каждого кластера; создана программа, выполняющая валидацию качества N и G генов в конкретной пробе и вычисляющая мутации, позволяющие идентифицировать штамм вируса бешенства. Данные исследования дают представление о разнообразии вируса бешенства, а также могут помочь проследить пути его эволюции и распространения, что важно не только для научных исследований, но и для практического применения при эпидемиологическом анализе на различных территориях.

ЛИТЕРАТУРА

1. Sadeuh-Mba S.A., Momo J.B., Besong L., Loul S., Njouom R. Molecular characterization and phylogenetic relatedness of dog-derived Rabies Viruses circulating in Cameroon between 2010 and 2016. *PLoS Negl Trop Dis*. 2017;11(10):e0006041. DOI: 10.1371/journal.pntd.0006041.
2. Reddy G.M., Singh R., Singh R.P., Singh K.P., Gupta P.K., Desai A., Verma R. Molecular characterization of Indian rabies virus isolates by partial sequencing of nucleoprotein (N) and phosphoprotein (P) genes. *Virus Genes*. 2011;43(1):13–17. DOI: 10.1007/s11262-011-0601-0.
3. Ботвинкин А.Д., Зарва И.Д., Якович Н.В., Адельшин Р.В., Мельникова О.В., Андаев Е.И., Щепин С.Г. Эпидемиологический анализ вспышек бешенства в Забайкалье после трансграничного заноса инфекции. *Эпидемиология и инфекционные болезни. Актуальные вопросы*. 2019;9(3):15–24. DOI: 10.18565/epidem.2019.9.3.15-24.
4. Гулюкин А.М., Шабейкин А.А., Макаров В.В., Зайкова О.Н., Гребенникова Т.В., Забережный А.Д., Южаков А.Г. Особенности эпизоотологического процесса и молекулярно-генетическая характеристика изолятов вируса бешенства, выявленных на территории Тверской области. *Вопросы вирусологии*. 2018;63(3):115–123. DOI: 10.18821/0507-4088-2018-63-3-115-123.
5. Kuzmin I.V., Botvinkin A.D., McElhinney L.M., Smith J.S., Orciari L.A., Hughes G.J., Rupprecht C.E. Molecular epidemiology of terrestrial rabies in the former Soviet Union. *J Wildl Dis*. 2004;40(4):617–631. DOI: 10.7589/0090-3558-40.4.617.
6. Девяткин А.А., Лукашев А.Н., Полещук Е.М., Ткачёв С.Е., Дедков В.Г., Сидоров Г.Н., Шипулин Г.А. Молекулярная эпидемиология вируса бешенства на территории Российской Федерации. *Эпидемиология и вакцинопрофилактика*. 2017;16(1):39–42. DOI: 10.31631/2073-3046-2017-16-1-39-42.
7. Troupin C., Dacheux L., Tanguy M., Sabeta C., Blanc H., Bouchier C., Bourhy H. Large-scale phylogenomic analysis reveals the complex evolutionary history of rabies virus in multiple carnivore hosts. *PLoS Pathog*. 2016;12(12):e1006041. DOI: 10.1371/journal.ppat.1006041.

8. Nextstrain: New Resources for Rabies, Lassa, and Yellow Fever Virus [Internet]. Available from: <https://nextstrain.org/blog/2025-02-21-rabies-lassa-yellow-fever-resources>. Accessed April 17, 2025.
9. Campbell K., Gifford R.J., Singer J., Hill V., O'toole A., Rambaut A., Brunker K. Making genomic surveillance deliver: A lineage classification and nomenclature system to inform rabies elimination. PLoS Pathog. 2022;18(5):e1010023. DOI: 10.1371/journal.ppat.1010023.
10. Rambaut A., Holmes E.C., O'Toole Á., Hill V., McCrone J.T., Ruis C., Pybus O.G. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020;5(11):1403–1407. DOI: 10.1038/s41564-020-0770-5.
11. Talbi C., Lemey P., Suchard M.A., Abdelatif E., Elharrak M., Jalal N., Bourhy H. Phylodynamics and human-mediated dispersal of a zoonotic virus. PLoS pathogens. 2010;6(10):e1001166. DOI: 10.1371/journal.ppat.1001166

© Герасименко Артем Александрович (gerasimenko_aa@antiplague.ru); Горох Алевтина Михайловна;
Писанов Руслан Вячеславович; Водопьянов Алексей Сергеевич
Журнал «Современная наука: актуальные проблемы теории и практики»

ИССЛЕДОВАНИЕ ТОКСИКОЛОГИЧЕСКИХ ХАРАКТЕРИСТИК ТЕРМИЧЕСКИ ОБРАБОТАННОГО ОСАДКА СТОЧНЫХ ВОД

STUDY OF THE TOXICOLOGICAL CHARACTERISTICS OF THE THERMONECHANICALLY TREATED SEWAGE SLUDGE

T. Kalinnikova
A. Egorova
A. Gatiyatullina
T. Kuznetsova
I. Knyazev
A. Petrov

Summary. Under laboratory conditions the influence of water extract from thermally treated urban sewage sludge on survivability of small crustaceans, holotrichous ciliates, soil nematodes as well as on the growth of higher plants was studied. The water extract did not have toxic effect on infusoria *Paramecium caudatum*. Acute toxic effect of water extract on cladocerans *Ceriodaphnia affinis* was eliminated by 2.5-fold dilution. In experiments with soil nematodes *Caenorhabditis elegans* the extract in concentration range 4–100 % caused the death of more than 70 % of animals. The adverse effect of water extract on spring wheat *Triticum vulgare* was eliminated by 1.2-fold dilution. Results of testing on aquatic organisms confirm that the examined thermally treated urban sewage sludge may be classified as hazard class IV according its environmental impact level; and it is non-toxic for monocotyledonous agricultural plants.

Keywords: sewage sludges, toxicity, *Caenorhabditis elegans*, *Paramecium caudatum*, *Ceriodaphnia affinis*, *Triticum vulgare*.

Калинникова Татьяна Борисовна

кандидат биологических наук, зав. лабораторией,
Институт проблем экологии и недропользования
Академии наук Республики Татарстан, г. Казань
tbkalinnikova@gmail.com

Егорова Анастасия Васильевна

младший научный сотрудник, Институт проблем
экологии и недропользования Академии наук
Республики Татарстан, г. Казань
egorovanastassia@gmail.com

Гатиятуллина Алсу Фоатовна

младший научный сотрудник, Институт проблем
экологии и недропользования Академии наук
Республики Татарстан, г. Казань
gaf9212@gmail.com

Кузнецова Татьяна Васильевна

научный сотрудник, Институт проблем экологии
и недропользования Академии наук
Республики Татарстан, г. Казань
kuznetsovatyana@mail.ru

Князев Игорь Владимирович

научный сотрудник, Институт проблем экологии
и недропользования Академии наук
Республики Татарстан, г. Казань
Kneze3@yandex.ru

Петров Андрей Михайлович

кандидат биологических наук, зав. лабораторией,
Институт проблем экологии и недропользования
Академии наук Республики Татарстан, г. Казань
zepam2@rambler.ru

Аннотация. В лабораторных условиях изучено влияние водной вытяжки из термически обработанного осадка городских сточных вод на выживаемость мелких рачков, равноресничных инфузорий, почвенных нематод и рост высших растений. Водная вытяжка не оказывала токсического действия на инфузорий *Paramecium caudatum*. Острое токсическое действие вытяжки на ветвистоусых рачков *Ceriodaphnia affinis* устранялось при 2.5-кратном разведении. В экспериментах с почвенными нематодами *Caenorhabditis elegans* вытяжка в концентрации 4–100 % вызывала гибель более 70 % особей. Вредное воздействие водной вытяжки на пшеницу яровую *Triticum vulgare* устранялось при разведении в 1.2 раза. Результаты тестирования на гидробионтах подтверждают, что исследованный термически обработанный осадок городских сточных вод по степени воздействия на окружающую среду относится к IV классу опасности (малотоксичная категория); для однодольных сельскохозяйственных растений он не токсичен.

Ключевые слова: осадки сточных вод, токсичность, *Caenorhabditis elegans*, *Paramecium caudatum*, *Ceriodaphnia affinis*, *Triticum vulgare*.

Введение

В процессе эксплуатации биологических городских очистных сооружений образуются значительные объёмы осадков сточных вод (ОСВ), включающих избыточный активный ил и осевшие в первичных отстойниках взвешенные вещества. Образующиеся осадки наряду с органическими и минеральными соединениями содержат условно патогенную и патогенную микрофлору, гельминтов и другие опасные для человека и окружающей среды компоненты, что требует их предварительной обработки перед утилизацией. В настоящее время на Казанских городских очистных сооружениях построен цех и внедрена технология обработки ОСВ предусматривающая их обезвоживание и сушку при температуре 105–110°C, которая обеспечивает элиминацию патогенной микрофлоры и других опасных для человека и окружающей среды живых организмов.

Таблица 1.

Характеристики гранулированных осадков сточных вод

Влажность, %	Плотность (ρ)	Орг. в-во, %	Зольность, %	N _{общ.} , %	P _{общ.} , %	pH _{сол.}
5–10	0.62	54.0–60.6	39.4–46.0	1.4–2.0	1.3–1.5	5.8–6.6

Образующийся после обработки устойчивый к влаге и механическому воздействию гранулированный продукт (ГП) содержит необходимые для развития растений и почвенных микроорганизмов органические вещества и биогенные элементы (табл. 1), что указывает на возможность его использования в качестве удобрения и для рекультивации деградированных и загрязнённых почв.

В 2022 году ООО «Бифар-Экология» ГП был сертифицирован в качестве «органического гранулированного удобрения на основе осадков сточных вод», что позволяет использовать его при рекультивации загрязнённых и деградированных почв [1].

Учитывая, то, что ГП содержит компоненты, например, тяжёлые металлы (табл. 2), которые в низких концентрациях могут оказывать комбинированное действие на биологические объекты, целесообразно определить токсическое действие ГП на почвенные организмы и растения. Поэтому целью настоящего исследования стало изучение влияния разных концентраций

водной вытяжки из термически обработанного осадка городских сточных вод на выживаемость мелких рачков, равноресничных инфузорий, нематод и рост высших растений.

Материалы и методы исследования

Объект исследования: гранулированный продукт, полученный в ходе термомеханической обработки осадка сточных вод очистных сооружений МУП «Водоканал» г. Казани.

Тест-объекты: Для определения токсичности полученного из осадка сточных вод ГП использовали батарею тест-объектов включающую равноресничных инфузорий *Paramecium caudatum* [2], ветвистоусых рачков *Ceriodaphnia affinis* [3], почвенных нематод *Caenorhabditis elegans* [4] и пшеницу яровую *Triticum vulgare* [5].

Подготовку водной вытяжки из ГП к анализам проводили в соответствии с [5] при массовом соотношении отхода и дистиллированной воды 1:10 [6].

Время тестирования на *Paramecium caudatum* составляло 24 часа, на *Ceriodaphnia affinis* — 48 часов, на *Triticum vulgare* — 96 часов при температуре 20–23°C и интенсивности освещения 4000 Лк в течение 12 часов в сутки. *Caenorhabditis elegans* культивировали в течение 24 часов при температуре 22°C.

Остальные условия тестирования соответствовали использованным методикам выполнения измерений [2–5]. Количество повторностей 3–5 в зависимости от тест-организма.

По результатам тестирования на животных определяли среднюю летальную концентрацию вызывающую гибель 50 % и более тест-организмов (LC₅₀), безвредную концентрацию вызывающую гибель не более 10 % гидробионтов (LC₁₀), не более 20 % червей (LC₂₀). При тестировании на пшенице устанавливали концентрацию ГП в водной вытяжке вызывающую 50 % (ER₅₀) и 20 % (ER₂₀) угнетение роста корней пшеницы.

Результаты и обсуждение

Использование в исследованиях широко применяемых в токсикологических исследованиях инфузорий

Таблица 2.

Содержание тяжёлых металлов в гранулированном продукте (средние значения)

Форма	Металл, мг/кг							
	Cd	Pb	Co	Cu	Ni	Zn	Cr	Mn
Валовая	4.47	42.20	7.70	111.10	50.10	428.60	167.30	125.40
Подвижная	0.23	2.20	1.59	0.55	12.10	102.40	3.12	85.90

P. caudatum и рачков *C. affinis* позволяет определить степень негативного воздействия (класс опасности) ГП на окружающую среду [6].

Известно, что паразитирующие на корнях растений нематоды, наносят большой ущерб растениеводству. Ежегодные потери урожая в мире вследствие поражения растений нематодами оцениваются в 12–25 % продукции [7], а в денежном выражении ущерб составляет 157 миллиардов евро в год [8]. Снижение наносимого нематодами растениям ущерба требует применения химических средств борьбы [9]. Применение пестицидов наносит ущерб окружающей среде, поскольку они уничтожают не только вредителей растений, но и других беспозвоночных. Кроме того, применяемые для борьбы с корневыми нематодами дозы пестицидов многократно превышают концентрации, применяемые для борьбы с другими беспозвоночными-вредителями [10], и могут быть опасны для конечных потребителей продукции растениеводства.

C. elegans является удобной моделью для изучения действия токсикантов на паразитических нематод. Как правило, вещества, токсичные для *C. elegans*, токсичны для паразитических нематод, и наоборот [11].

Компоненты ГП при его внесении в почву могут стимулировать или ингибировать рост растений. Для определения возможного влияния полученного ГП на рост сельскохозяйственных растений было проведено его токсикологическое тестирование на *T. vulgare*.

Проведенные исследования показали, что водная вытяжка из ГП в концентрации 1–100 % не приводит к гибели равноресничных инфузорий *P. caudatum*. При тестировании водной вытяжки на ветвистоусых рачках *C. affinis* острое токсическое действие (LC_{50}) устраняет-

ся при 2.5-кратном разведении, а вредное воздействие (LC_{10}) при разведении в 4.9 раза. Полученные результаты подтверждают, что по степени воздействия на окружающую среду ГП относится к IV классу опасности [6].

Согласно результатам исследований водная вытяжка из ГП в концентрации 4–100 % оказывает токсическое действие на *C. elegans*, вызывая за 24 ч гибель 73–100 % особей. При разведении в 50 раз (концентрация 2 %) доля погибших нематод снижается до 32 %, при разведении в 100 раз (1 %) гибель составляет 13 %. При более низких концентрациях вытяжки (0.1–0.5 %) смертность нематод не превышала 7 % (рис. 1). Проведенные расчеты показали, что острое токсическое действие (LC_{50}) на *C. elegans* устраняется при 44-кратном разведении водной вытяжки из ГП (концентрация 2.3 %), а вредное воздействие (LC_{20}) при разведении в 85 раз и концентрации ГП 1.2 % (рис. 1).

Тестирование водной вытяжки из ГП показало, что она не токсична для пшеницы, а её вредное воздействие на однодольное растение (ER_{20}) устраняется при разведении в 1.2 раза. Следует отметить, что при 50-кратном разведении вытяжки (концентрация 2 %) наблюдалась 22 % стимуляция роста корней пшеницы. В остальных испытанных вариантах присутствие ГП достоверно не влияло на развитие растений (рис. 2).

Заключение

Проведенные исследования показали, что ГП из осадка сточных вод не оказывает токсического действия на инфузорий *P. caudatum*. При тестировании на *C. affinis* острое токсическое действие водной вытяжки устраняется при 2.5-кратном разведении, а вредное воздействие (LC_{10}) при разведении в 4.9 раза.

Водная вытяжка из ГП при разведении до 25 (концентрация 4–100 %) оказывает токсическое действие на по-

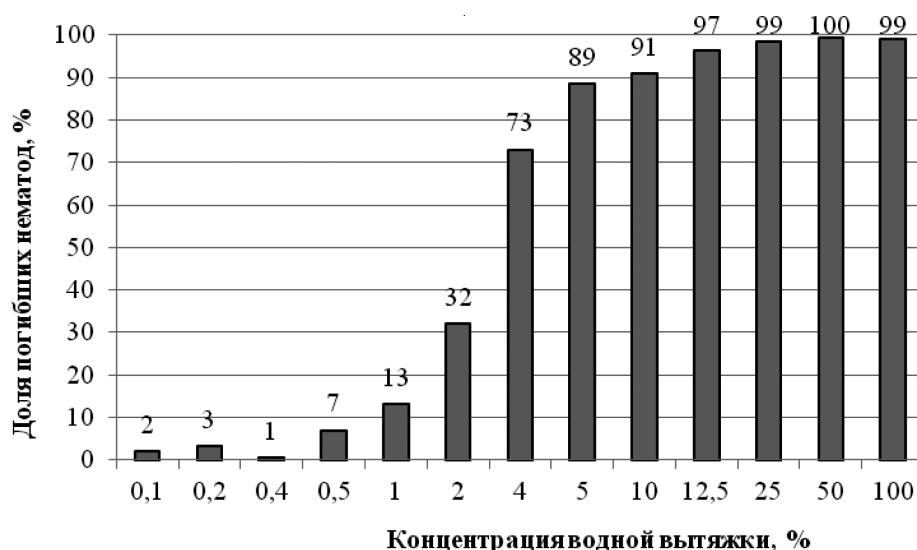


Рис. 1. Влияние водной вытяжки из ГП на выживаемость *Caenorhabditis elegans*

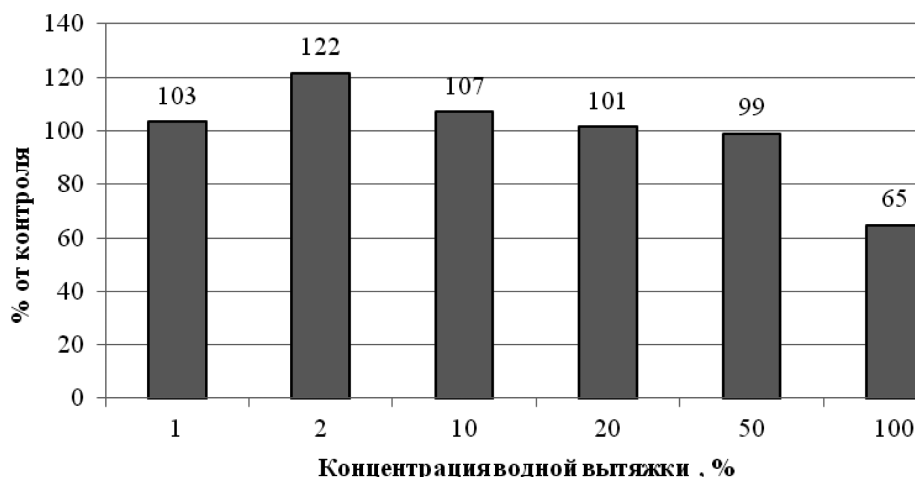


Рис. 2. Влияние водной вытяжки из ГП на рост корней пшеницы на ранней стадии развития
По оси ординат — длина проростков в % от контроля

чвенных нематод *C. elegans*, вызывая гибель более 70 % особей.

Вредное воздействие водной вытяжки на пшеницу (ER_{20}) устраняется при разведении в 1.2 раза.

Результаты тестирования на гидробионтах подтверждают, что ГП по степени воздействия на окружающую среду относится к IV классу опасности (малотоксичная категория), он не токсичен для однодольных сельскохозяйственных растений.

При значительном содержании паразитирующих нематод внесение ГП в почву приведет к снижению их численности.

По токсикологическим характеристикам, полученный в ходе термомеханической обработки осадка городских сточных вод ГП может быть использован при рекультивации загрязненных и деградированных почв, в случае необходимости снижения численности паразитических почвенных нематод при выращивании декоративных и иных растений.

ЛИТЕРАТУРА

1. Экологический сертификат соответствия № RA.RU.11HA15.П.00114 «Удобрение органическое гранулированное на основе осадков сточных вод, изготовляемое серийно по ТУ 37.00.20-001-03317648-2022 на очистных сооружениях канализации г. Казани». Выдано ООО «БИФАР-Экология».
2. ПНД Ф Т 14.1:2.3.13-06, ПНД Ф Т 16.1:2.3.3.10-06 Методика определения токсичности отходов, почв, осадков сточных, поверхностных и грунтовых вод методом биотестирования с использованием равноресничных инфузорий *Paramecium caudatum* Ehrenberg.
3. ФР 1.39.2007.03221 Методика определения токсичности воды и водных вытяжек из почв, осадков сточных вод, отходов по смертности и изменению плодовитости цериодафний. — Акварос, 2007.
4. ISO 10872. Water quality — Determination of the toxic effect of sediment and soil samples on growth, fertility and reproduction of *Caenorhabditis elegans* (Nematoda). First edition 2010-06-15. Reviewed and confirmed in 2015. Geneva: ISO Copyright office, 2010. 24 p.
5. МР 2.1.7.2297-07. Обоснование класса опасности отходов производства и потребления по фитотоксичности. 2.1.7. Почва. Очистка населенных мест. Бытовые и промышленные отходы. М., 2007. 15 с.
6. Приказ Минприроды России от 31.03.2025 N 158 «Об утверждении критериев отнесения отходов к I–V классам опасности по степени негативного воздействия на окружающую среду» (Зарегистрировано в Минюсте России 29.04.2025 N 82010).
7. Migunova V.D., Sasanelli N. Bacteria as biocontrol tool against phytoparasitic nematodes // Plants. 2021. V. 10. Article 389. doi: 10.3390/plants10020389
8. Ntalli N., Adamski Z., Doula M., Monokrousos N. Nematicidal amendments and soil remediation // Plants. 2020. V. 9. Article 429. doi: 10.3390/plants9040429
9. Ahmad Md. F., Ahmad F.A., Alsayegh A.A., Zeyaulah Md., AlSharani A.M., Muzammil K., Saati A.A., Wahab S., Elbendary E.Y., Kambal N., Abdelrahman M.H., Hussain S. Pesticides impact on human health and the environment with their mechanisms of action and possible countermeasures // Heliyon. 2024. Vol. 10. E29128. doi: 10.1016/j.heliyon.2024.e29128
10. Sasanelli N., Konrat A., Migunova V., Toderas I., Iurcu-Straistaru E., Rusu S., Bivol A., Andoni C., Veronico P. Review on control methods against plant parasitic nematodes applied in southern member states (C zone) of the European Union // Agriculture. 2021. Vol. 11. Article 602. doi: 10.3390/agriculture11070602
11. Калинникова Т.Б. Почвенная нематода *Caenorhabditis elegans* как модель для изучения паразитических нематод // Лабораторные животные для научных исследований. 2024. Т. 7, № 1. С. 61–68.

© Калинникова Татьяна Борисовна (tbkalinnikova@gmail.com); Егорова Анастасия Васильевна (egorovanastassia@gmail.com);

Гатиятуллина Алсу Фоатовна (gaf9212@gmail.com); Кузнецова Татьяна Васильевна (kuznetsovatatyana@mail.ru);

Князев Игорь Владимирович (Kneze3@yandex.ru); Петров Андрей Михайлович (zram2@rambler.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

ТВЕРДЫЕ ДИСПЕРСНЫЕ МАТЕРИАЛЫ КАК ЭКОЛОГИЧЕСКИ УСТОЙЧИВАЯ ФОРМА АККУМУЛИРОВАНИЯ ТЕПЛОТЫ

GRANULAR SOLID MATERIALS AS AN ENVIRONMENTALLY SUSTAINABLE FORM OF HEAT STORAGE

**E. Markin
M. Terekhova
E. Otrubyannikov**

Summary. The article substantiates the application of granular solid materials of natural and industrial origin as an environmentally safe medium for heat storage in energy systems with a high share of renewable generation. Physical mechanisms of heat transfer in packed beds are examined, including conductive, convective, and radiative components, along with the influence of porosity, particle shape, and particle size distribution on the effective thermal conductivity of the layer. Based on a comparative analysis of quartz sand, basalt-group rocks, aluminosilicate slags, and ceramic granules, it is shown that synthetic granules possess the highest volumetric heat capacity, while medium-fraction slag materials provide an optimal balance between thermodynamic characteristics and environmental acceptability. A nonlinear dependence between the intensification of interphase heat transfer and the growth of pressure drop is identified, allowing determination of a narrow range of energy-efficient operating conditions. Integration of experimental data and life cycle assessment confirms the viability of using industrial mineral wastes in regenerative thermal energy storage systems.

Keywords: heat storage, packed bed, granular solid materials, heat transfer, aluminosilicate slags, life cycle assessment, environmental sustainability.

Маркин Евгений Михайлович

ФГБОУ ВО «РГУ им. А.Н. Косыгина»

zhenya.markin.00@gmail.com

Терехова Мария Владимировна

ФГБОУ ВО «РГУ им. А.Н. Косыгина»

terekhova-mv@rguk.ru

Отрубянников Егор Владимирович

Кандидат технических наук, доцент

ФГБОУ ВО «РГУ им. А.Н. Косыгина»

otrubyannikov-ev@rguk.ru

Аннотация. В статье обосновано применение твердых дисперсных материалов природного и техногенного происхождения в качестве экологически безопасной среды для аккумулирования теплоты в энергосистемах с высокой долей возобновляемой генерации. Рассмотрены физические механизмы теплопереноса в зернистых засыпках, включающие кондуктивную, конвективную и радиационную составляющие, а также влияние порозности, формы частиц и гранулометрического состава на эффективную теплопроводность слоя. На основе сравнительного анализа кварцевого песка, горных пород базальтовой группы, алюмосиликатных шлаков и керамических гранул показано, что синтетические гранулы обладают наибольшей объемной теплоемкостью, тогда как среднефракционные шлаковые материалы обеспечивают оптимальный баланс между термодинамическими характеристиками и экологической приемлемостью. Выявлена нелинейная зависимость между интенсификацией межфазного теплообмена и ростом аэродинамических потерь давления, позволяющая определить узкий диапазон энергетически эффективных режимов эксплуатации. Интеграция экспериментальных данных и оценки жизненного цикла подтверждает перспективность использования техногенных минеральных отходов в регенеративных теплоаккумуляторах.

Ключевые слова: аккумулирование теплоты, зернистый слой, твердые дисперсные материалы, теплоперенос, алюмосиликатные шлаки, оценка жизненного цикла, экология.

Введение

Глобальная декарбонизация энергетики обуславливает стремительный рост доли нестабильной возобновляемой генерации, что предъявляет принципиально новые требования к системам буферного аккумулирования тепловой энергии. Систематический обзор технологий аккумулирования теплоты в зернистом слое [1, с. 634] свидетельствует о том, что зернистые засыпки из природных и техногенных минеральных материалов представляют одну из наиболее перспективных платформ для высокотемпературного аккумулирования, сочетая структурную простоту, термическую стабильность и экологическую безопасность.

Принципы аккумулирования теплоты находят применение в гражданском строительстве, где латентное теплонакопление с использованием фазопереходных материалов позволяет стабилизировать температурный режим ограждающих конструкций [2, с. 61]; перспективным направлением является также комбинирование sensible и латентного теплонакопления в единой установке промышленного масштаба [3, с. 662]. В отличие от систем на расплавленных солях или синтетических маслах, твердые дисперсные среды не подвержены термической деградации с выделением токсичных летучих фракций и способны устойчиво функционировать при температурах до 1000 °C [4, с. 3090; 5, с. 1780].

Фундаментальная физика теплопереноса в зернистых слоях описывается суперпозицией кондуктивного, конвективного и радиационного механизмов, каждый из которых чувствителен к микроструктурным параметрам засыпки [6, с. 934; 7, с. 190]; систематизация закономерностей теплообмена и гидромеханики в дисперсных сквозных потоках выполнена в работе [8, с. 424]. Классические корреляции для аэродинамического сопротивления [9, с. 91] и для межфазной теплоотдачи [10, с. 326; 11, с. 364] заложили аналитический фундамент моделирования аккумуляторов теплоты с зернистым слоем, однако перенос этих моделей на системы с техногенными шлаковыми засыпками требует учета специфической морфологии и химической неоднородности частиц. Проблема осложняется тем, что уменьшение размера гранул, интенсифицирующее теплообмен, одновременно резко увеличивает потери давления и энергоемкость подготовки сырья, формируя многокритериальную оптимизационную задачу [12, с. 814].

Экологическая парадигма современной энергетики требует вовлечения вторичных материальных ресурсов в технологический оборот. Металлургические шлаки, золы-уноса тепловых электростанций и отвальные породы горно-обогатительных комбинатов составляют гигантский резервуар алюмосиликатных и магнезиальных соединений с высокой удельной энтальпией [13, с. 780; 14, с. 608]. Исследования в рамках европейского проекта REslag продемонстрировали возможность промышленной эксплуатации аккумуляторов теплоты с зернистым слоем со шлаковой засыпкой для утилизации тепла отходящих газов электродуговой печи [15, с. 710], а оценки жизненного цикла показали существенное преимущество отхода-ориентированных систем перед традиционными аналогами по индикаторам климатического воздействия и водопотребления [16; 17, с. 1512].

Целью настоящей работы является комплексное обоснование применения твердых дисперсных материалов различного генезиса в качестве теплоаккумулирующих сред, реализуемое посредством интеграции термофизического, гидродинамического и экологического анализа с привлечением верифицированных численных моделей.

Материалы и методы исследования

Инструментальная база исследования включала комплексный термографический и гидродинамический анализ четырех групп дисперсных материалов: природных кварцевых песков, дробленого габбро-диабазы, алюмосиликатных шлаков электродуговых печей и синтезированных керамических гранул на основе спеченного боксита. Выбор материалов обоснован системным анализом кандидатных теплоаккумулирующих сред [18, с. 427; 19, с. 117], критериями которого выступали термическая стабильность в диапазоне до 800 °C, химическая инерт-

ность, доступность и стоимость. Экспериментальная характеристика шлаковых фракций проводилась по методологии, апробированной в работе [14, с. 608], включающей определение плотности, удельной теплоемкости и теплопроводности при температурах до 1000 °C.

Подготовка образцов включала предварительную сушку при 105 °C до постоянной массы ($\Delta m < 0,1\%$) и фракционирование на стандартных ситах с размером ячеек 1, 5, 10 и 20 мм. Гранулометрический состав каждой фракции контролировался лазерной дифракцией (коэффициент вариации диаметра частиц не превышал 12 %). Порозность засыпки определялась объемно-массовым методом путем заливки мерным объемом воды при пятикратном повторении укладки; среднее значение порозности составило $\epsilon = 0,38 \pm 0,02$ для шлаковых гранул и $\epsilon = 0,41 \pm 0,02$ для дробленого габбро-диабазы. Шероховатость поверхности частиц оценивалась по оптическим микрофотографиям; шлаковые гранулы характеризовались выраженной нерегулярной морфологией с коэффициентом сферичности $0,72 \pm 0,06$ по Вадделу, тогда как керамические гранулы отличались высокой сферичностью ($0,94 \pm 0,03$). Все измерения термофизических свойств проводились не менее чем в трех повторностях; в табл. 1 представлены средние значения, а систематическая погрешность определения теплоемкости методом дифференциальной сканирующей калориметрии не превышала $\pm 3\%$, теплопроводности методом лазерной вспышки — $\pm 5\%$, плотности пикнометрическим методом — $\pm 0,5\%$.

Определение изобарной теплоемкости и коэффициента температуропроводности осуществлялось методом дифференциальной сканирующей калориметрии при скорости нагрева не более 5 К/мин в атмосфере аргона. Установлено, что температурная зависимость теплоемкости твердого тела оказывает доминирующее влияние на динамику заряда аккумулятора [5, с. 1802], что потребовало проведения измерений в широком температурном диапазоне. Для оценки термоциклической стабильности материалы подвергались многократному нагреву-охлаждению, при этом кварцсодержащие образцы демонстрировали деградацию при α - β фазовом переходе кварца (~ 573 °C), тогда как керамические и шлаковые гранулы сохраняли исходные свойства [20, с. 43].

Газодинамическое сопротивление зернистого слоя исследовалось на аэродинамическом стенде замкнутого типа, оснащенном прецизионными датчиками дифференциального давления (класс точности 0,25 %, калибровка по эталонному микроманометру перед каждой серией опытов) и термоанемометрами, обеспечивающими измерения в диапазоне чисел Рейнольдса от 100 до 5000. Каждый режим повторялся трижды; неопределенность измерения перепада давления составляла $\pm 4\%$, коэффициента теплоотдачи — $\pm 7\%$ (оценка

по методу Клайна–Макклитока). Обработка данных базировалась на уравнении [9, с. 92], комбинирующем вязкостный член Козени–Кармана и инерционный член Бёрка–Пламмера. Вычислительная часть работы основывалась на двухфазной модели локальной термической неравновесности пористой среды [21, с. 1270], реализованной в рамках подхода вычислительной гидродинамики с осесимметричной геометрией и температурозависимыми свойствами фаз. Расчетная область представляла собой цилиндрический аккумулятор высотой 1,2 м и внутренним диаметром 0,6 м; использовалась неструктурированная тетраэдрическая сетка из приблизительно 185 тыс. элементов со сгущением вблизи стенок ($y^+ < 5$). Независимость решения от сетки подтверждена сериями расчетов на трех последовательно измельченных сетках (92, 185 и 370 тыс. элементов): максимальное отклонение температурного поля между двумя последними сетками не превышало 0,3 %. Граничные условия включали равномерный профиль скорости и температуры на входе ($T_{вх} = 800$ °C, $v_{вх} = 0,35$ м/с), условие свободного истечения на выходе и адиабатность боковой стенки; начальная температура засыпки принималась равной 25 °C. Интегрирование по времени выполнялось неявной схемой второго порядка с адаптивным шагом (начальный $\Delta t = 0,1$ с, максимальный $\Delta t = 1$ с); критерий сходимости на каждом временном шаге — снижение невязок уравнений энергии и импульса до 10^{-6} .

Потери эксергии в табл. 2 определялись как отношение необратимых потерь к полной эксергии подведенного теплового потока: $\eta_{ex} = 1 - (E_{out}/E_{in})$, где $E_{in} = Q_{in} \cdot (1 - T_0/T_{вх})$ — эксергия входящего потока при температуре окружающей среды $T_0 = 298$ K, а $E_{out} = Q_{out} \cdot (1 - T_0/T_{вых})$ — эксергия выходящего потока. Указанная метрика учитывает как термодинамическое обесценивание теплоты при контакте с засыпкой, так и механические потери на нагнетание газа; последние рассчитывались по произведению объемного расхода на измеренный перепад давления.

Эффективная теплопроводность стационарного слоя рассчитывалась по ячеечной модели [6, с. 935; 7, с. 190] с привлечением справочных данных по теплопроводности пористых и зернистых сред [22, с. 264], учитывающей порозность засыпки, отношение теплопроводностей фаз, контактную проводимость и радиационный теплообмен между поверхностями частиц. Корреляция [10, с. 820] $Nu = 2 + 1,1 \cdot Pr^{1/3} \cdot Re^{0,6}$ использовалась для определения конвективной теплоотдачи газ-твердое тело в диапазоне $Re = 15-8500$.

Экологическая оценка основывалась на методологии анализа жизненного цикла (ОЖЦ) с использованием пяти индикаторов воздействия: потенциал глобального потепления (ПГП), совокупное потребление энергии (СПЭ), истощение абиотических ресурсов, фор-

мирование твердых частиц и эвтрофикация [16]. Учитывались прямые затраты энергии на процессы дробления и фракционирования, выбросы мелкодисперсной пыли при транспортировке, а также потенциальная способность материалов к выщелачиванию тяжелых металлов. Валидационная база данных включала результаты более 400 независимых экспериментов, опубликованных в профильных рецензируемых журналах за последние два десятилетия [1, с. 630; 23, с. 1065].

Результаты и обсуждение

Термофизические характеристики исследуемых материалов

Термодинамическая емкость зернистого слоя определяется сочетанием физических свойств индивидуальных частиц и макроскопической структуры упаковки. Скрининг кандидатных материалов для высокотемпературного аккумулирования теплоты [18, с. 250; 19, с. 119] показывает, что наиболее перспективными являются породы базальтовой группы (долерит, габбро), характеризующиеся высокой механической прочностью и термостабильностью, а также керамические гранулы из спеченного боксита, обладающие выдающейся циклоустойкостью [20, с. 45]. Результаты экспериментального определения базовых термофизических параметров представлены в табл. 1.

Таблица 1.

Базовые термофизические параметры исследуемых дисперсных материалов

Материал	Истинная плотность, кг/м ³	Удельная теплоемкость, Дж/(кг·К)	Эффективная теплопроводность, Вт/(м·К)	Объемная теплоемкость, кДж/(м ³ ·К)
Кварцевый песок	2645	835	1,14	2209
Габбро-диабаз	2988	792	1,87	2367
Алюмосиликатный шлак (EAF)	2812	916	1,35	2576
Керамические гранулы (боксит)	3155	1045	2,12	3297

Примечание к табл. 1: истинная плотность, удельная теплоемкость (при 200 °C) и теплопроводность определены авторами экспериментально (методами гелиевой пикнометрии, дифференциальной сканирующей калориметрии и лазерной вспышки соответственно). Объемная теплоемкость вычислена как произведение плотности и удельной теплоемкости. Приведенные значения представляют средние по трем независимым измерени-

ям; соответствующие погрешности указаны в разделе «Материалы и методы». Табл. 2 содержит результаты собственных аэродинамических экспериментов с относительной погрешностью измерения перепада давления $\pm 4\%$ и коэффициента теплоотдачи $\pm 7\%$; эксергетические потери рассчитаны по формуле, приведенной там же. Табл. 3 объединяет собственные результаты помола и удельных выбросов; углеродный эквивалент пересчитан по коэффициентам IPCC 2021 для регионального энергобаланса. Табл. 4 сопоставляет расчетные данные с экспериментальными значениями пилотной установки [4, с. 3088]; доверительный интервал экспериментальных температур составляет $\pm 2,5^\circ\text{C}$ (термопары типа К, класс 1).

Анализ табличных данных выявляет, что керамические гранулы на основе спеченного боксита превосходят природные материалы по всем ключевым параметрам (рис. 2а), включая объемную теплоемкость — интегральный показатель, определяющий массогабаритные характеристики теплообменного аппарата. Данный результат согласуется с экспериментальной характеристикой гранулированных сред [20, с. 44], подтвердившей полную термоциклическую стабильность керамики при температурах до 800°C . Шлаки электродуговой печи (ЕАФ) занимают промежуточное положение: при удельной теплоемкости $916\text{ Дж}/(\text{кг}\cdot\text{K})$ они обеспечивают объемную теплоемкость $2576\text{ кДж}/(\text{м}^3\cdot\text{K})$, что согласуется с данными работы [14, с. 607], зафиксировавшей теплопроводность шлаков порядка $1,5\text{--}1,7\text{ Вт}/(\text{м}\cdot\text{K})$.

Кварцевый песок, несмотря на широкую доступность, демонстрирует наименьшую объемную теплоемкость и, что критически важно, подвержен структурной деградации при фазовом переходе $\alpha\text{--}\beta$ кварца при 573°C [20, с. 46; 24, с. 181]. Габбро-диабаз характеризуется высокой теплопроводностью ($1,87\text{ Вт}/(\text{м}\cdot\text{K})$), однако пониженная удельная теплоемкость обусловлена доминированием тяжелых металлических оксидов в минералогическом составе, обладающих низким фоновым спектром колебаний кристаллической решетки. Эффективная теплопроводность засыпки определяется не только свойствами индивидуальных частиц, но и контактным взаимодействием, которое описывается моделью [6, с. 940; 7, с. 200] с учетом коэффициента сплющивания и эффекта Смолуховского.

Динамика фильтрационного теплообмена и аэродинамическое сопротивление

Переход от статического анализа к изучению динамических характеристик системы требует оценки аэродинамических потерь при нагнетании газообразного теплоносителя сквозь зернистый слой. Возникает фундаментальная инженерная дилемма: уменьшение диаметра гранул экспоненциально увеличивает коэффици-

ент межфазного теплообмена, сокращая время заряда аккумулятора, но одновременно провоцирует нелинейный рост перепада давления, описываемый уравнением [9, с. 91]. Результаты экспериментального исследования динамических характеристик межфазного теплообмена для среднефракционного шлакового наполнителя представлены в табл. 2.

Таблица 2.
Динамические характеристики межфазного теплообмена в шлаковой засыпке

Скорость фильтрации, м/с	Перепад давления, Па/м	Коэффициент теплоотдачи, $\text{Вт}/(\text{м}^2\cdot\text{K})$	Потеря эксергии, %
0,15	125	46	2,1
0,35	458	89	5,7
0,65	1346	142	14,9
1,05	3128	199	31,4

Функциональный анализ зависимости перепада давления от скорости фильтрации подтверждает переход от ламинарного к развитому турбулентному режиму течения в поровых каналах. Разложение экспериментальных данных показывает, что показатель степени при переменной скорости отклоняется от линейного закона Дарси и приближается к квадратичной форме Форхгеймера, что свидетельствует о доминировании инерционных сил в порах сложной геометрии (рис. 1). Данный результат полностью согласуется с классическим уравнением [9, с. 93], объединяющим вязкостный и инерционный вклады в единую корреляцию, а также с отечественными расчетными рекомендациями для стационарных зернистых слоев [25, с. 176], и с корреляциями для теплоотдачи между слоем и погруженными поверхностями, обобщенными в фундаментальной монографии по псевдоожигу [26, с. 664].

Приращение коэффициента теплоотдачи подчиняется логарифмическому закону с выраженным асимптотическим торможением при скоростях свыше $0,5\text{ м/с}$ (рис. 1). Анализ табличных данных выявляет, что керамические гранулы на основе спеченного боксита превосходят природные материалы по всем ключевым параметрам, что соответствует корреляции [10, с. 329]. Расчет вторых производных функции теплоотдачи выявляет точку перегиба, после которой энергетические затраты на нагнетание газа начинают превалировать над приращением интенсивности теплосъема. Эксергетические потери при форсированных режимах возрастают до $31,4\%$, что свидетельствует о значительной генерации энтропии в пограничном слое. Статистическая обработка методом наименьших квадратов позволила аппроксимировать поверхность отклика, оптимум которой локализован в узком диапазоне скоростей $0,25\text{--}0,45\text{ м/с}$,

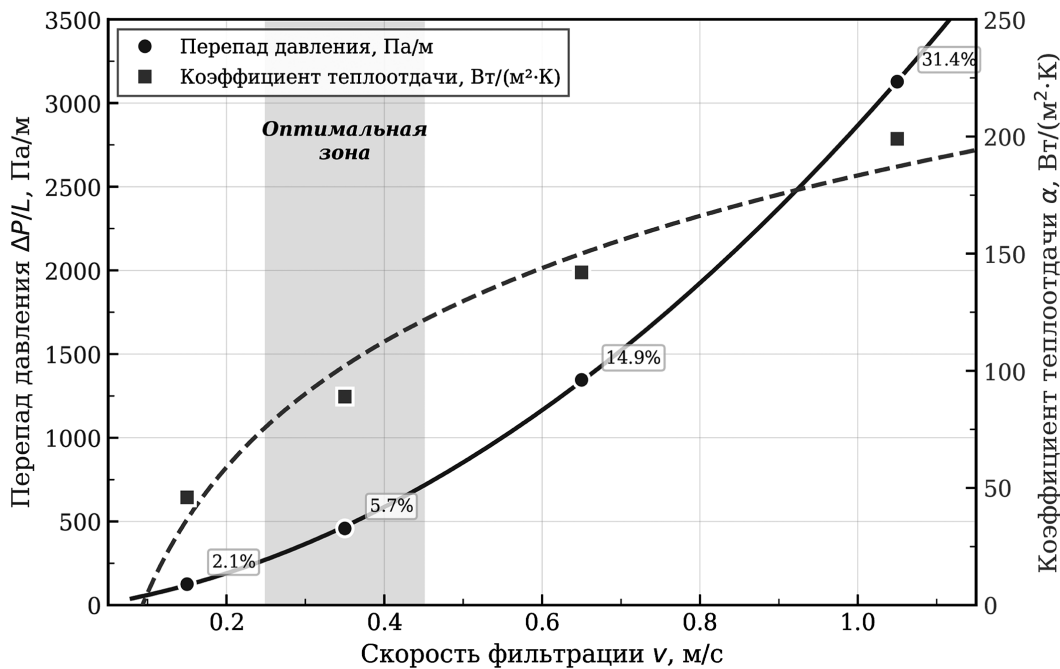


Рис. 1. Зависимость перепада давления и коэффициента теплоотдачи от скорости фильтрации в шлаковой засыпке

где отношение переданного теплового потока к затраченной мощности достигает локального максимума.

Экологическая оценка и анализ жизненного цикла

Экологическая парадигма требует интеграции показателей физической эффективности с метриками антропогенного воздействия на биосферу. Подготовка дисперсных материалов сопряжена со значительными энергозатратами на механическое разрушение горной породы или шлакового монолита. Результаты анализа жизненного цикла [16; 17, с. 1512] показывают, что совокупный экологический след аккумулирующей системы определяется не только прямой эмиссией CO₂ при производстве электроэнергии, но и потенциалом истощения абиотических ресурсов и индексами экотоксичности (табл. 3).

Таблица 3.

Показатели экологического воздействия на этапах жизненного цикла

Фракция материала	Затраты энергии на помол, кВт·ч/т	Углеродный эквивалент, кг CO ₂ /т	Индекс экотоксичности, отн. ед.
Крупная (>20 мм)	45	12,9	0,014
Средняя (5–20 мм)	129	36,4	0,028
Мелкая (1–5 мм)	345	98,2	0,057
Микродисперсная (<1 мм)	892	253	0,124

Энергетические затраты на измельчение демонстрируют степенную зависимость от целевого размера фрак-

ции, согласующуюся с классическими законами теории измельчения. Уменьшение медианного диаметра частиц вдвое приводит к нелинейному скачку потребления энергии вследствие перехода от хрупкого разрушения макродефектов к пластической деформации микроструктуры. В исследовании [16] установлено, что системы аккумулирования теплоты на основе вторичной керамики из спеченной золы-уноса при 500 °С обеспечивают ПГП порядка 1,2 кг CO₂-экв/кВт·ч при энергетической окупаемости менее трех месяцев и КЭОИ (коэффициент энергетической отдачи инвестиций) порядка 55, что вдвое превышает показатели природного газа. В работе [17, с. 1514] продемонстрировано, что системы аккумулирования на основе отходов снижают климатическое воздействие на 40 % и водопотребление на 60 % по сравнению с двухбаковой системой на расплавленных солях.

Индекс экотоксичности обнаруживает тревожную тенденцию увеличения химической активности частиц по мере роста удельной поверхности (рис. 26). Экспоненциальное возрастание площади контакта с атмосферной влагой интенсифицирует процессы экстракции тяжелых металлов из аморфной матрицы шлаков, что устанавливает жесткий нижний предел допустимого размера гранул для открытых систем терморегулирования. Это ограничение, наряду с углеродной нагрузкой микроизмельчения, формирует зону запрета в пространстве проектных решений, эффективно исключая микродисперсные порошки из арсенала экологически ответственного проектирования, несмотря на их превосходные теплообменные характеристики.

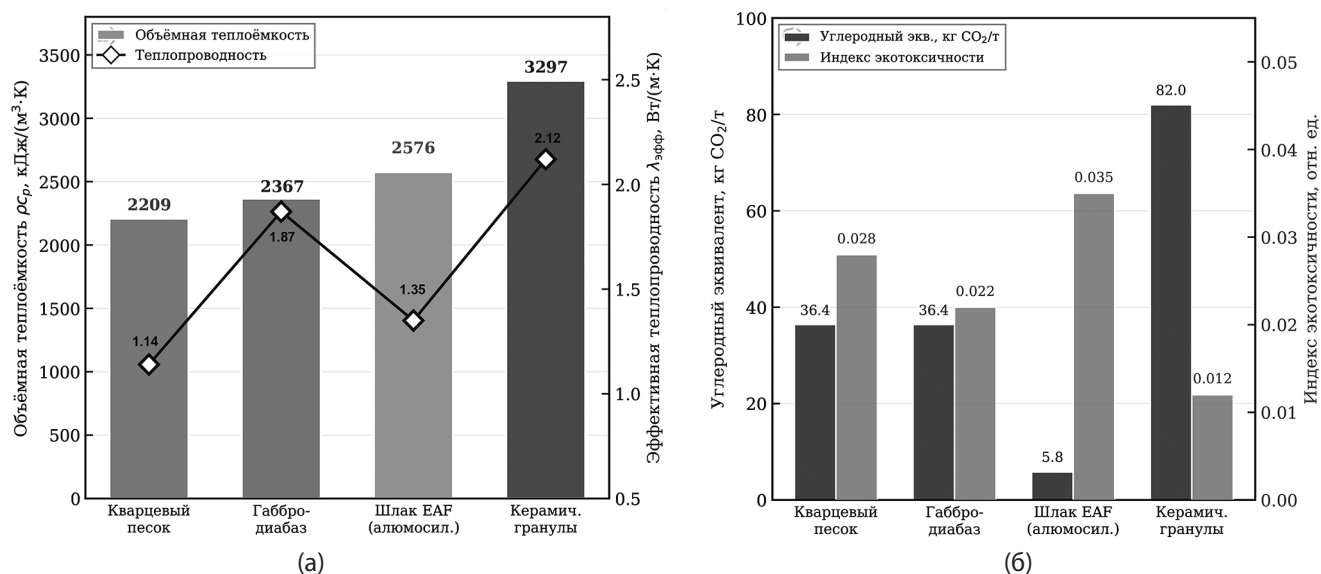


Рис. 2. Сравнительные характеристики дисперсных материалов: а) термофизические свойства; б) экологические показатели (средняя фракция 5–20 мм)

Верификация вычислительной модели

Валидация численной модели осуществлялась посредством сопоставления расчетных полей температур с экспериментальными данными пилотных установок [4, с. 3087; 5, с. 1799; 21, с. 1270]. Двухфазная модель с учетом локальной термической неравновесности между твердым скелетом и газовым потоком реализована в осесимметричной геометрии с неструктурированными тетраэдрическими сетками, сгущенными в зонах контактного взаимодействия гранул. Результаты верификации температурного фронта при полном цикле заряда представлены в табл. 4.

Таблица 4.

Верификация численной модели: расчетные и экспериментальные значения температуры

Глубина слоя, м	$T_{\text{эксперимент}}, ^\circ\text{C}$	$T_{\text{расчет}}, ^\circ\text{C}$	Относительная погрешность, %
0,25	785	782	0,4
0,50	543	552	1,6
0,75	313	298	4,6
1,00	125	115	8,4

Пространственный анализ погрешностей обнаруживает закономерное нарастание расхождения по мере удаления от входного сечения аккумулятора. В лобовой зоне (глубина 0,25 м), где температурные градиенты сглажены и тепловое равновесие между фазами наступает быстро, погрешность минимальна (0,4 %). Однако на фронте тепловой волны (глубина 1,0 м) накапливается вычислительная ошибка, достигающая 8,4 %, что обусловлено алгоритмической жесткостью схемы ин-

тегрирования при описании резких скачков энтальпии. Аналогичная тенденция зафиксирована в верификационных исследованиях для засыпок из дробленого шпата [5, с. 1801] и каменных пород [4, с. 3096]. Модель склонна предсказывать более крутой температурный градиент, чем наблюдается экспериментально — явление численной дисперсии, корректируемое введением искусственных диссипативных членов [21, с. 1269].

Многокритериальная оптимизация

Интеграция термофизических, гидродинамических и экологических массивов данных формирует многомерное пространство параметров, в котором осуществляется поиск глобального оптимума. Построение фронта Парето для конфликтующих целевых функций — минимизации аэродинамического сопротивления, минимизации углеродного следа и максимизации объемной теплоемкости — выявляет узкий коридор допустимых решений. Топологический анализ пространства решений показывает, что использование среднефракционных (5–20 мм) алюмосиликатных шлаков с узким гранулометрическим распределением позволяет достичь баланса между интенсификацией теплообмена и приемлемым уровнем энергозатрат на подготовку сырья.

Формально задача многокритериальной оптимизации формулировалась следующим образом. Вектор варьируемых параметров включал медианный диаметр частиц d (1–25 мм), скорость фильтрации v (0,1–1,2 м/с) и тип материала засыпки (дискретный параметр). Три целевые функции определялись как: $f_1(d, v) = \Delta P/L \rightarrow \min$ (удельные аэродинамические потери по уравнению [9, с. 90]); $f_2(d) = E_{\text{помол}}(d) \cdot k_{\text{CO}_2} \rightarrow \min$ (углеродный след подготовки материала, кг CO₂-экв/т); $f_3(\text{материал}) = \rho \cdot c_p \rightarrow \max$ (объемная теплоемкость засыпки). Ограничения задачи

включали: индекс экотоксичности $I_{tox} \leq 0,05$ отн. ед., потерю эксергии $\eta_{ex} \leq 15\%$, термоциклическую стабильность (менее 2 % деградации теплоемкости за 500 циклов) и минимальную скорость теплосъема $Q/(m \cdot \Delta T) \geq 50$ Вт/(м³·К). Фронт Парето строился методом NSGA-II (размер популяции 200 особей, 500 поколений, вероятность скрещивания 0,9, вероятность мутации 0,1); критерием остановки служила стабилизация гиперобъема фронта Парето (изменение менее 0,1 % за последние 50 поколений). Штрафные функции вводились мультипликативно: при превышении I_{tox} значение целевых функций f_1 и f_2 умножалось на коэффициент $(1 + 100 \cdot (I_{tox} - 0,05)^2)$, что эффективно вытесняло недопустимые решения из популяции.

Введение штрафных функций за превышение индекса экотоксичности радикально трансформирует форму поверхности отклика, отсекая области применения микродисперсных порошков. Данный результат согласуется с рекомендациями систематических обзоров [13, с. 764], каталогизировавших отходы промышленности как кандидатные теплоаккумулирующие материалы с учетом экономических и экологических преимуществ перед традиционными средами. Расчет собственных значений гесса в точке оптимума подтверждает устойчивость найденной конфигурации к флуктуациям входных параметров, что гарантирует стабильную работу системы в условиях стохастического изменения режимов зарядки. Промышленный опыт проекта [15, с. 715] подтверждает теоретические выводы, демонстрируя техническую осуществимость аккумулятора теплоты с зернистым слоем со шлаковой засыпкой на действующем металлургическом предприятии.

Заключение

Синтез термофизических закономерностей и экологических императивов формирует новую парадигму проектирования систем аккумулирования теплоты на базе твердых дисперсных сред. Экспериментально и расчетно доказана возможность создания высокоемких регенеративных матриц, превосходящих традицион-

ные теплоносители по критериям термодинамической стабильности, безопасности и экологической приемлемости. Керамические гранулы из спеченного боксита обеспечивают наивысшую объемную теплоемкость (3297 кДж/(м³·К)), однако среднефракционные алюмосиликатные шлаки электродуговых печей представляют наиболее рациональный компромисс, сочетая удовлетворительные термофизические показатели с минимальным экологическим следом и практически нулевой стоимостью сырья.

Установлен оптимальный диапазон скоростей фильтрации газового теплоносителя (0,25–0,45 м/с), в котором отношение переданного теплового потока к затраченной мощности нагнетания достигает максимума при допустимом уровне эксергетических потерь. Показано, что уменьшение размера частиц ниже 5 мм приводит к непропорциональному росту углеродного следа и индекса экотоксичности, что исключает микродисперсные фракции из экологически обоснованных проектных решений.

Верификация двухфазной численной модели локальной термической неравновесности по экспериментальным данным пилотных установок подтвердила адекватность прогнозирования температурного поля с погрешностью, не превышающей 8,4 % на фронте тепловой волны, что соответствует инженерным требованиям к проектированию промышленных теплоаккумуляторов. Применение техногенных минеральных образований в энергетическом секторе решает двойную задачу: обеспечивает надежный инструмент сглаживания пиковых нагрузок и способствует масштабной утилизации промышленных отходов, снижая антропогенную нагрузку на литосферу. Дальнейшее развитие технологии связано с созданием интеллектуальных систем предиктивного управления режимами заряда-разряда и гармонизацией нормативно-правовой базы для перевода вторичных минеральных ресурсов в статус сертифицированной продукции энергетического назначения.

ЛИТЕРАТУРА

- Esence T., Bruch A., Molina S., Stutz B., Fourmigué J.-F. A review on experience feedback and numerical modeling of packed-bed thermal energy storage systems // *Solar Energy*. — 2017. — Vol. 153. — P. 628–654. DOI: 10.1016/j.solener.2017.03.032
- Геллер Ю.А. Изменение температуры внутренней поверхности массивных конструкций зданий при использовании фазопереходных материалов // *Промышленные процессы и технологии*. — 2024. — Т. 4. — № 4(14). — С. 54–64. DOI: 10.37816/2713-0789-2024-4-4(14)-54-64
- Geissbühler L., Kolman M., Zanganeh G., Haselbacher A., Steinfeld A. Analysis of industrial-scale high-temperature combined sensible/latent thermal energy storage // *Applied Thermal Engineering*. — 2016. — Vol. 101. — P. 657–668. DOI: 10.1016/j.applthermaleng.2015.12.031
- Zanganeh G., Pedretti A., Zavattoni S.A., Barbato M.C., Steinfeld A. Packed-bed thermal storage for concentrated solar power — Pilot-scale demonstration and industrial-scale design // *Solar Energy*. — 2012. — Vol. 86. — P. 3084–3098. DOI: 10.1016/j.solener.2012.07.019
- Hänchen M., Brückner S., Steinfeld A. High-temperature thermal storage using a packed bed of rocks — Heat transfer analysis and experimental validation // *Applied Thermal Engineering*. — 2011. — Vol. 31. — P. 1798–1806. DOI: 10.1016/j.applthermaleng.2010.10.034
- Zehner P., Schlünder E.U. Wärmeleitfähigkeit von Schüttungen bei mäßigen Temperaturen // *Chemie Ingenieur Technik*. — 1970. — Bd. 42. — S. 933–941. DOI: 10.1002/cite.330421408

7. Bauer R., Schlünder E.U. Effective radial thermal conductivity of packings in gas flow. Part II: Thermal conductivity of the packing fraction without gas flow // *International Chemical Engineering*. — 1978. — Vol. 18(2). — P. 189–204.
8. Горбис З.Р. Теплообмен и гидромеханика дисперсных сквозных потоков. — М.: Энергия, 1970. — 424 с.
9. Ergun S. Fluid flow through packed columns // *Chemical Engineering Progress*. — 1952. — Vol. 48. — P. 89–94.
10. Wakao N., Kaguei S., Funazkri T. Effect of fluid dispersion coefficients on particle-to-fluid heat transfer coefficients in packed beds: Correlation of Nusselt numbers // *Chemical Engineering Science*. — 1979. — Vol. 34(3). — P. 325–336. DOI: 10.1016/0009-2509(79)85064-2
11. Wakao N., Kaguei S. Heat and Mass Transfer in Packed Beds (Topics in Chemical Engineering, Vol. 1). — New York: Gordon and Breach, 1982. — 364 p.
12. Zanganeh G., Pedretti A., Haselbacher A., Steinfeld A. Design of packed bed thermal energy storage systems for high-temperature industrial process heat // *Applied Energy*. — 2015. — Vol. 137. — P. 812–822. DOI: 10.1016/j.apenergy.2014.07.110
13. Gutierrez A., Miró L., Gil A., Rodríguez-Aseguinolaza J., Barreneche C., Calvet N., Py X., Fernández A.I., Grágeda M., Ushak S., Cabeza L.F. Advances in the valorization of waste and by-product materials as thermal energy storage (TES) materials // *Renewable and Sustainable Energy Reviews*. — 2016. — Vol. 59. — P. 763–783. DOI: 10.1016/j.rser.2015.12.071
14. Ortega-Fernández I., Calvet N., Gil A., Rodríguez-Aseguinolaza J., Faik A., D'Aguanno B. Thermophysical characterization of a by-product from the steel industry to be used as a sustainable and low-cost thermal energy storage material // *Energy*. — 2015. — Vol. 89. — P. 601–609. DOI: 10.1016/j.energy.2015.05.153
15. Ortega-Fernández I., Rodríguez-Aseguinolaza J. Thermal energy storage for waste heat recovery in the steelworks: The case study of the REslag project // *Applied Energy*. — 2019. — Vol. 237. — P. 708–719. DOI: 10.1016/j.apenergy.2019.01.007
16. Lalau Y., Al Asmi I., Olivès R., Dejean G., Meffre A., Py X. Energy analysis and life cycle assessment of a thermal energy storage unit involving conventional or recycled storage materials and devoted to industrial waste heat valorisation // *Journal of Cleaner Production*. — 2022. — Vol. 330. — Art. 129950. DOI: 10.1016/j.jclepro.2021.129950
17. Lalau Y., Py X., Meffre A., Olivès R. Comparative LCA between current and alternative waste-based TES for CSP // *Waste and Biomass Valorization*. — 2016. — Vol. 7(6). — P. 1509–1519. DOI: 10.1007/s12649-016-9549-6
18. Tiskatine R., Oaddi R., Ait El Cadi R., Bazgaou A., Bouirden L., Aharoune A., Ihlal A. Suitability and characteristics of rocks for sensible heat storage in CSP plants // *Solar Energy Materials and Solar Cells*. — 2017. — Vol. 169. — P. 245–257. DOI: 10.1016/j.solmat.2017.05.033
19. Khare S., Dell'Amico M., Knight C., McGarry S. Selection of materials for high temperature sensible energy storage // *Solar Energy Materials and Solar Cells*. — 2013. — Vol. 115. — P. 114–122. DOI: 10.1016/j.solmat.2013.03.009
20. Baumann T., Zunft S. Properties of granular materials as heat transfer and storage medium in CSP application // *Solar Energy Materials and Solar Cells*. — 2015. — Vol. 143. — P. 38–47. DOI: 10.1016/j.solmat.2015.06.037
21. Cascetta M., Cau G., Puddu P., Serra F. A comparison between CFD simulation and experimental investigation of a packed-bed thermal energy storage system // *Applied Thermal Engineering*. — 2016. — Vol. 98. — P. 1263–1272. DOI: 10.1016/j.applthermaleng.2016.01.019
22. Дульнев Г.Н., Заричняк Ю.П. Теплопроводность смесей и композиционных материалов: Справочная книга. — Л.: Энергия, 1974. — 264 с.
23. Singh H., Saini R.P., Saini J.S. A review on packed bed solar energy storage systems // *Renewable and Sustainable Energy Reviews*. — 2010. — Vol. 14(3). — P. 1059–1069. DOI: 10.1016/j.rser.2009.10.022
24. Allen K.G., von Backström T.W., Kröger D.G., Kisters A.F.M. Rock bed storage for solar thermal power plants: Rock characteristics, suitability, and availability // *Solar Energy Materials and Solar Cells*. — 2014. — Vol. 126. — P. 170–183. DOI: 10.1016/j.solmat.2014.03.030
25. Аэров М.Э., Тодес О.М., Наринский Д.А. Аппараты со стационарным зернистым слоем. Гидравлические и тепловые основы работы. — Л.: Химия, 1979. — 176 с.
26. Гельперин Н.И., Айнштейн В.Г., Кваша В.Б. Основы техники псевдоожижения. — М.: Химия, 1967. — 664 с.

© Маркин Евгений Михайлович (zhenya.markin.00@gmail.com); Терехова Мария Владимировна (terekhova-mv@rguk.ru);

Отрубянников Егор Владимирович (otrubbyannikov-ev@rguk.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

КОНЦЕПЦИЯ «ЕДИНОЕ ЗДОРОВЬЕ» (ONE HEALTH) В СИСТЕМЕ ЭКОЛОГИЧЕСКОЙ БЕЗОПАСНОСТИ

THE ONE HEALTH CONCEPT IN THE ENVIRONMENTAL SAFETY SYSTEM

S. Trunova
S. Nurmagomedova
U. Magomedova
R. Daudova
R. Suleymanova

Summary. The article examines the One Health concept as an interdisciplinary approach to ensuring environmental safety under contemporary global challenges. The relevance of the study is обусловлена the increasing interconnection between environmental conditions, animal health, and public health, which necessitates a shift from isolated management decisions to integrated models. The aim of the study is to analyze the potential of the One Health concept and to assess the possibilities of its implementation within the environmental safety system. Through the generalization of environmental and epidemiological data, relationships between anthropogenic pressure, ecosystem transformation, and the growth of environmentally обусловленные and zoonotic risks have been identified. It has been established that climate change and biodiversity loss contribute to the expansion of disease distribution areas and the emergence of new risk zones. The results demonstrate that the implementation of One Health principles enhances the effectiveness of environmental safety systems by integrating monitoring, interagency cooperation, and preventive risk management strategies. At the same time, several limitations have been identified, including institutional fragmentation and insufficient development of regulatory and methodological frameworks.

Keywords: One Health concept, environmental safety, zoonotic diseases, environment, public health, biodiversity, anthropogenic pressure, environmental monitoring.

Трунова Саният Акаевна

Кандидат биологических наук, доцент,
Дагестанский государственный медицинский
университет (Махачкала)
saniatakaeva@mail.ru

Нурмагомедова Саният Гаджиевна

Кандидат медицинских наук, доцент,
Дагестанский государственный медицинский
университет (Махачкала)
nurma.san@yandex.ru

Магомедова Умият Абдулбасировна

Старший преподаватель,
Дагестанский государственный медицинский
университет (Махачкала)
m.umijat@yandex.ru

Даудова Рабият Далгатовна

Кандидат биологических наук, доцент,
Дагестанский государственный педагогический
университет им Р. Гамзатова (Махачкала)
medicina_19@mail.ru

Сулейманова Раиса Герейхановна

Кандидат медицинских наук, доцент,
Дагестанский государственный медицинский
университет (Махачкала)
suleimanova.r.g@mail.ru

Аннотация. Концепция «Единое здоровье» (One Health) рассматривается как междисциплинарный подход к обеспечению экологической безопасности в условиях современных глобальных вызовов. Актуальность исследования обусловлена усилением взаимосвязи между состоянием окружающей среды, здоровьем животных и общественным здоровьем, что требует перехода от изолированных управленческих решений к интеграционным моделям. Целью работы является анализ потенциала концепции «Единое здоровье» и оценка возможностей ее внедрения в систему экологической безопасности. На основе обобщения данных экологического и эпидемиологического характера выявлены взаимосвязи между уровнем антропогенной нагрузки, трансформацией экосистем и ростом экологически обусловленных и зоонозных рисков. Установлено, что изменение климатических условий и снижение биоразнообразия способствуют расширению ареалов распространения заболеваний и формированию новых зон риска. Результаты исследования показывают, что внедрение принципов «Единого здоровья» позволяет повысить эффективность системы экологической безопасности за счет интеграции мониторинга, межведомственного взаимодействия и перехода к превентивным стратегиям управления рисками. Вместе с тем выявлены ограничения, связанные с институциональной разобщенностью и недостаточной разработанностью нормативно-методической базы.

Ключевые слова: концепция «Единое здоровье», экологическая безопасность, зоонозные заболевания, окружающая среда, общественное здоровье, биоразнообразие, антропогенная нагрузка, экологический мониторинг.

Процесс урбанизации, изменение климата, деградация экосистем и утрата биоразнообразия приводят к возникновению проблем, затрагивающих не только природу, но и здоровье человека и животных. Традиционные подходы к решению экологических и санитарно-эпидемиологических проблем оказываются недостаточно эффективными. Особую актуальность приобретает концепция «Единое здоровье» (One Health), характеризующаяся усилиями в области медицины, ветеринарии, экологии и других смежных дисциплинах [8, с. 16]. Данный подход исходит из того, что здоровье человека связано с состоянием окружающей среды и здоровьем животных, а возникающие угрозы, такие как зоонозные инфекции, загрязнение окружающей среды и изменение климата, требуют координированных решений [4, с. 7]. Включение концепции в систему экологической безопасности позволяет по-новому взглянуть на механизмы предупреждения и минимизации экологических рисков. Речь идет о мониторинге состояния окружающей среды и о формировании моделей взаимодействия между природными и антропогенными системами.

Цель нашего исследования — анализ потенциала внедрения концепции «Единое здоровье» в программу экологической безопасности, выявление ключевых направлений ее интеграции в существующие управленческие практики.

Формирование концепции «Единое здоровье» связано с осознанием взаимозависимости здоровья человека, животных и окружающей среды, что нашло отражение в ряду научных подходов второй половины XX — начала XXI века [5, с. 116]. Первоначально проблематика рассматривалась в рамках ветеринарии и изучения зоонозных инфекций, однако со временем исследовательский фокус расширился на экологические и социальные аспекты [2]. Деградация природных систем напрямую влияет на эпидемиологическую ситуацию, усиливая распространение инфекционных заболеваний, в том числе передающихся от животных к человеку [6, с. 17].

В ряде исследований активно используется термин «Экологическое здоровье» (EcoHealth), которое акцентирует внимание на устойчивости экосистем и их способности поддерживать устойчивость экосистем и их способности поддерживать жизнедеятельность биологических и социальных систем. Обе модели — «Единое здоровье» и «Экологическое здоровье» во многом пересекаются, однако первая концепция в большей степени ориентирована на экосистемные процессы, тогда как вторая — на интеграцию медицинских и ветеринарных аспектов с экологическими факторами [14, с. 149]. Кроме того, существует концепция «Планетарное здоровье» (Planetary Health), которая рассматривает здоровье человека на макроуровне — изменение климата в пла-

нетарных масштабах, глобальные биогеохимические циклы [10, с. 24]. Все три подхода формируют единую научную платформу, направленную на переосмысление взаимодействия человека и окружающей среды.

Одной из основных проблем, по мнению ученых, является проблемы продовольственной безопасности. Качество пищевых продуктов напрямую связано с здоровьем человека, благополучием животных и экологией [7, с. 149]. Акцентируется внимание на усилении ветеринарного контроля, экологическом мониторинге и межведомственном взаимодействии. Именно аспект продовольственной безопасности должен рассматриваться ключевым в концепции «Единое здоровье».

В рамках рассматриваемой модели уделяется внимание мониторингу и оценке экологических и санитарных рисков [13, с. 116]. Их уровень определяется природными и социально-экономическими условиями (уровень доходов населения; доступ к медицинским услугам; качество инфраструктуры). Вовлеченность населения способствует повышению эффективности профилактических мер и укреплению экологической устойчивости [9].

Эффективность внедрения концепции зависит от уровня координации, доступности данных и развития механизмов обмена информацией между различными ведомствами [12]. В ряде работ подчеркивается необходимость адаптации универсальных методик к локальным условиям [13, с. 116]. Различия в терминологии, методах и приоритетах могут затруднить интеграцию [8, с. 16].

В последние годы существенно возрос интерес к цифровым и аналитическим инструментам, применяемым в рамках концепции «Единое здоровье» — технология Big Data, геоинформационные системы, технологии дистанционного зондирования для отслеживания экологических изменений и распространения заболеваний [11]. Интеграция разнородных массивов данных позволяет выявлять скрытые закономерности и повысить точность прогнозирования рисков. Превентивные меры, направленные на поддержание устойчивости экосистем и раннее выявление угроз, в долгосрочной перспективе оказываются более эффективными по сравнению с затратами на ликвидацию последствий [9]. Существующие же правовые механизмы зачастую фрагментарны и не обеспечивают комплексного подхода к проблемам [3].

Концепция «Единое здоровье» постепенно трансформируется из узкоспециализированного направления в комплексную научно-практическую парадигму, ориентированную на обеспечение экологической безопасности. Методологическую основу исследования составили подходы, позволяющие рассмотреть взаимосвязи состояния окружающей среды, здоровья человека и животных. Исследование основывается на данных о состоянии

окружающей среды и распространении заболеваемости среди фауны [2; 3; 9; 11; 12].

В условиях высокой антропогенной нагрузки наблюдается рост показателей в частоте контактов человека с дикой фауной, уровне загрязнения воздуха. Одновременно фиксируется снижение биоразнообразия и деградация экосистем (Таблица 1).

Таблица 1.

Влияние экологических факторов на показатели риска для здоровья (в рамках концепции «Единое здоровье»)

Показатель	Группа территорий с низкой антропогенной нагрузкой	Группа территорий с высокой антропогенной нагрузкой	Изменение (%)
Частота зоонозных заболеваний (на 100 тыс. чел.)	18	37	+105,6 %
Уровень биоразнообразия (индекс, усл. ед.)	0,78	0,42	-46,2 %
Частота контактов человека с дикой фауной (случ./год)	12	29	+141,7 %
Распространённость инфекций среди животных (%)	9,5	21,3	+124,2 %
Загрязнение воздуха (PM _{2.5} , мкг/куб. м.)	12	34	+183,3 %
Доля экологически обусловленных заболеваний (%)	14,2	28,7	+102,1 %
Темпы изменения ареалов переносчиков (%)	6,8	15,9	+133,8 %

Внедрение единых принципов охраны здоровья способствует повышению эффективности системы экологической безопасности путем интеграции данных и координации действий различных субъектов. Сочетание экологического и эпидемиологического мониторинга позволяет лучше прогнозировать риски и оптимизировать распределение требуемых ресурсов. Концепция «Единое здоровье», которая является практическим и целенаправленным подходом к обеспечению экологической безопасности, направлена на предотвращение возможных рисков, а не на улучшение традиционных моделей. Она определяет стратегию профилактики контроля заболеваний. Во время пандемии COVID-19 накопленный опыт развитых стран в области медицинского и экологического мониторинга помог избежать больших жертв. В то же время развивающиеся государства оказались более подвержены риску заболеваний зооноза из-за ограниченности ресурсов или слабых институтов. Универсальные (традиционные) модели реагирования на чрезвычайные ситуации в области здравоохранения и экологии не всегда эффективно могут предсказать риски или решить насущные проблемы.

В качестве выводов мы можем указать следующее:

- концепция «Единое здоровье» отражает взаимосвязь состояния окружающей среды, здоровья животных и человека;
- применение принципов «Единого здоровья» повышает эффективность мониторинга и управления рисками за счет междисциплинарного и межведомственного взаимодействия;
- существующие институциональные механизмы ограничены;
- адаптация концепции «Единое здоровье» к региональным условиям необходима с учетом природно-климатических и социально-экономических факторов;
- целесообразно развитие нормативно-правовой базы и цифровых инструментов.

ЛИТЕРАТУРА

1. Калининкова, Т.Б. Подход «Единое здоровье» как возможный способ оценки экологических рисков (на примере цианотоксинов) / Т.Б. Калининкова, А.В. Егорова, Р.Р. Шагидуллин // Труды Карельского научного центра Российской академии наук. — 2021. — № 9. — С. 130–145. — DOI 10.17076/lim1362. — EDN BUZHJG.
2. Концепция «Единое здоровье»: [Электронный ресурс]. М., 2025. URL: <https://www.who.int/ru/news-room/fact-sheets/detail/one-health> (Дата обращения: 18.03.2026).
3. Концепция «Единое здоровье»: комплексный подход к здоровью человека, животных и окружающей среды: [Электронный ресурс]. М., 2025. URL: https://www.wipo.int/ru/web/global-health/w/news/2023/news_0009 (Дата обращения: 18.03.2026).
4. Концепция «Единое здоровье»: комплексный подход к здоровью человека, животных и окружающей среды / А.В. Концевая, А.А. Анциферова, Д.К. Муканеева [и др.] // Профилактическая медицина. — 2024. — Т. 27, № 11. — С. 7–12. — DOI 10.17116/profmed2024271117. — EDN CLOAKA.
5. Макаров, В.В. «Единое здоровье» / В.В. Макаров, А.А. Стекольников, А.С. Донченко // Сибирский вестник сельскохозяйственной науки. — 2025. — Т. 55, № 11(324). — С. 116–128. — DOI 10.26898/0370-8799-2025-11-12. — EDN ELBZDG.
6. Маммедова, А. Концепция «Единое здоровье» (One Health): взаимосвязь здоровья животных, человека и окружающей среды / А. Маммедова, К. Сахетмурадов // Новая наука: от идеи к результату. — 2026. — № 2. — С. 17–19. — EDN PIJJJC.
7. Панин, А.Н. Безопасность продовольствия как основа концепции «ЕДИНОЕ ЗДОРОВЬЕ» / А.Н. Панин // Труды Всероссийского НИИ экспериментальной ветеринарии имени Я.П. Коваленко. — 2021. — Т. 82, № 1. — С. 149. — EDN CZLXIC.

8. Подход «Единое здоровье» (One Health) (обзорная статья) / А.М. Затевалов, В.И. Гашенко, Н.В. Гудова, О.Г. Гречишникова // Биотехнология в медицине и фармации. — 2025. — Т. 2, № 1. — С. 16–25. — DOI 10.51620/3034-7211-2025-1-2-16-25. — EDN TSFGRE.
9. Представлен совместный план действий «Единое здоровье» для устранения угроз здоровью людей, животных, растений и окружающей среды: [Электронный ресурс]. М., 2025. URL: <https://www.who.int/ru/news/item/17-10-2022-one-health-joint-plan-of-action-launched-to-address-health-threats-to-humans-animals-plants-and-environment> (Дата обращения: 18.03.2026).
10. Пути к планетарному и общественному здоровью / Г. Клаус, Д. Майк, Й. Никлас, К. Алексия // Современная школа России. Вопросы модернизации. — 2025. — № 3(70). — С. 24–27. — EDN WAMLMA.
11. Пять вещей, которые нужно знать о концепции «Единое здоровье» в Европейском регионе ВОЗ: [Электронный ресурс]. М., 2025. URL: <https://www.who.int/europe/ru/news/item/23-03-2023-5-things-to-know-about-one-health-in-the-who-european-region> (Дата обращения: 18.03.2026).
12. Совместный план действий ФАО, ЮНЕП, ВОЗ и МЭБ: [Электронный ресурс]. М., 2025. URL: <https://ecfs.msu.ru/news/sovместnyij-plan-dejstvij-fao-yunep-voz-i-meb> (Дата обращения: 18.03.2026).
13. Awareness of the «One Health» principles among specialists of the public Health system and veterinary service of Kazakhstan / T. Zh. Akpanova, T.K. Rakhypbekov, A.K. Mussakhanova [et al.] // Science & Healthcare. — 2024. — Vol. 26, No. 3. — P. 116–122. — DOI 10.34689/sh.2024.26.3.013. — EDN HDWQGX.
14. Huang, Zh. Ecological security: the insurmountable line of defense for human health / Zh. Huang // Вестник МАНЭБ. — 2025. — Vol. 30, No. 4. — P. 149–156. — EDN EDYDUY.

© Трунова Саният Акаевна (saniatakaeva@mail.ru); Нурмагомедова Саният Гаджиевна (nurma.san@yandex.ru);
 Магомедова Умият Абдулбасировна (m.umijat@yandex.ru); Даудова Рабият Далгатовна (medicina_19@mail.ru);
 Сулейманова Раиса Герейхановна (suleimanova.r.g@mail.ru)
 Журнал «Современная наука: актуальные проблемы теории и практики»

МЕТОДИКА ФУНКЦИОНАЛЬНОЙ ЭМУЛЯЦИИ ПРОТОКОЛА NEW IP В СРЕДЕ GNS3 И ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА ЭФФЕКТИВНОСТИ ИНКАПСУЛЯЦИИ ДЛЯ IOT-СЕГМЕНТОВ ПЕРСПЕКТИВНЫХ СЕТЕЙ

A METHODOLOGY FOR FUNCTIONAL EMULATION OF THE NEW IP PROTOCOL IN THE GNS3 ENVIRONMENT AND AN EXPERIMENTAL EVALUATION OF ENCAPSULATION EFFICIENCY FOR IOT SEGMENTS OF PROMISING NETWORKS

**N. Adaev
N. Pozhidaev
P. Markaryan
M. Taranets
S. Malakhov**

Summary. The New IP protocol, proposed by Huawei and considered by ITU-T SG13 as an element of the 2030 network architecture, does not have a commercial software implementation, significantly hindering its independent experimental evaluation. In this paper, we propose and test a methodology for functional emulation of the New IP protocol using publicly available tools: GNS3, Docker, and Python 3.11. It has been analytically proven that the variable-length New IP header (12 bytes for the IoT segment) reduces encapsulation overhead by 70.0 % compared to the fixed 40-byte IPv6 header, increasing link efficiency from 16.7 % to 40.0 % with an 8-byte payload. Experiments using correct UDP echo measurements of real RTT and DSCP EF (RFC 3246) marking for both protocols revealed that, under identical QoS conditions, the network behavior of New IP and IPv6 is statistically indistinguishable: when sampling 1000 critical packets, the loss is 0 % for both protocols with active discard of ~187,845 background packets per session, confirming the 100 % effectiveness of the DSCP EF mechanism for both protocols. The competitive advantage of New IP for promising IoT networks lies in its bandwidth savings.

Keywords: New IP; IPv6; GNS3; IoT; encapsulation; DSCP EF; QoS; functional emulation; 6G networks.

Адаев Николай Олегович

Самарский национальный исследовательский университет имени академика С.П. Королева
adaev.kol@yandex.ru

Пожидает Николай Андреевич

Самарский национальный исследовательский университет имени академика С.П. Королева
pzhidaev63nik@gmail.com

Маркарян Петрос Габриелович

Самарский национальный исследовательский университет имени академика С.П. Королева
markaryan.petros@yandex.ru

Таранец Максим Евгеньевич

Самарский национальный исследовательский университет имени академика С.П. Королева
mtaranetz@yandex.ru

Малахов Сергей Валерьевич

к.т.н., доцент,
Самарский национальный исследовательский университет имени академика С.П. Королева
malakhov.sv@ssau.ru

Аннотация. Протокол New IP, предложенный компанией Huawei и рассматриваемый в ITU-T SG13 как элемент архитектуры сетей 2030, не имеет коммерческой программной реализации, что существенно затрудняет его независимую экспериментальную оценку. В настоящей работе предложена и апробирована методика функциональной эмуляции New IP на базе общедоступных инструментов — GNS3, Docker и Python 3.11. Аналитически доказано, что заголовок переменной длины New IP (12 байт для IoT-сегмента) снижает накладные расходы на инкапсуляцию на 70,0 % по сравнению с фиксированным 40-байтным заголовком IPv6, увеличивая КПД канала с 16,7 % до 40,0 % при полезной нагрузке 8 байт. Экспериментально, с использованием корректного UDP echo-измерения реального RTT и маркировки DSCP EF (RFC 3246) для обоих протоколов, установлено, что при одинаковых условиях QoS сетевое поведение New IP и IPv6 статистически неразличимо: при выборке 1000 критических пакетов потери составляют 0 % для обоих протоколов при активном отбросе ~187 845 фоновых пакетов за сеанс, что подтверждает стопроцентную эффективность механизма DSCP EF для обоих протоколов. Конкурентное преимущество New IP для перспективных IoT-сетей лежит в плоскости экономии полосы пропускания.

Ключевые слова: New IP; IPv6; GNS3; IoT; инкапсуляция; DSCP EF; QoS; функциональная эмуляция; сети 6G.

Развитие сетей шестого поколения, промышленного интернета вещей и систем управления критической инфраструктурой предъявляет принципиально новые требования к сетевому уровню: минимальные накладные расходы при инкапсуляции коротких сообщений, детерминированная доставка с гарантированными параметрами задержки и надёжности [1,2]. Существующий стек IPv6, проектировавшийся для универсальной передачи данных по принципу «наилучших усилий», не в полной мере удовлетворяет этим требованиям: фиксированный заголовок длиной 40 байт создаёт неприемлемо высокие накладные расходы при передаче типичных IoT-сообщений, а именно 4–8 байт полезной нагрузки.

Протокол New IP представляет собой концептуальную архитектуру сетевого уровня, внесённую компанией Huawei на рассмотрение Международного союза электросвязи (ITU-T SG13) в рамках программы «Сети 2030» [3]. Независимый анализ предложения New IP, проведённый ICANN, выявил ряд архитектурных неточностей и несогласованностей в исходном документе Huawei [4]. Ключевыми отличиями New IP являются: адресация переменной длины от 1 до 128 байт, механизм «сетевых контрактов» для резервирования ресурсов с гарантированными параметрами качества обслуживания и семантически-ориентированная маршрутизация. Вместе с тем ни одна коммерческая операционная система не реализует стек New IP, что существенно ограничивает возможности его объективной экспериментальной оценки.

Цель работы — разработать воспроизводимую методику функциональной эмуляции New IP и экспериментально проверить две гипотезы: гипотеза 1 о снижении накладных расходов на инкапсуляцию благодаря заголовку переменной длины New IP; гипотеза 2 об отсутствии статистически значимой разницы в сетевом поведении New IP и IPv6 при равных условиях QoS [5].

Стандарт IPv6 определяет базовый заголовок фиксированной длины 40 байт, включающий 16-байтные поля адресов источника и назначения. При полезной нагрузке 8 байт, например в типичное IoT-сообщении, коэффициент накладных расходов составляет 83,3 %. Это неприемлемо для узкополосных каналов LoRaWAN и NB-IoT с полосой 0,3–50 кбит/с [1,2].

New IP предлагает три принципиальных улучшения по сравнению с IPv6. Во-первых, адресация переменной длины: для локального IoT-сегмента достаточно 1–2 байт адреса, расчётный заголовок составляет 8–12 байт [3]. Во-вторых, механизм сетевых контрактов: отправитель согласует с сетью параметры потока, маршрутизаторы обрабатывают пакеты с контрактной меткой детерминированно [4]. В-третьих, семантически-ориентированная маршрутизация позволяет передавать в заголовке требования к качеству обслуживания наряду с адресом назначения.

Поскольку программный стек New IP отсутствует, в настоящей работе применяется методология функциональной эмуляции: размер заголовка определяется аналитически на основе спецификаций ITU-T [3,4], а сетевое поведение воспроизводится программно. Для обеспечения корректности сравнения используется стандартный механизм DSCP EF применительно к обоим протоколам, что функционально соответствует приоритизации, аналогичной сетевым контрактам New IP, но реализованной стандартными средствами IPv4/IPv6 DiffServ [6,8].

Коэффициент полезного использования канала (КПД инкапсуляции):

$$\eta = L_p / (L_z + L_p) \quad (1)$$

где L_p — полезная нагрузка (байт), L_z — длина заголовка (байт).

Коэффициент накладных расходов:

$$\varphi = L_z / (L_z + L_p) \quad (2)$$

Относительное снижение накладных расходов New IP по сравнению с IPv6:

$$\Delta\varphi = (\varphi_{IPv6} - \varphi_{NewIP}) / \varphi_{IPv6} \times 100 \% \quad (3)$$

Для IoT-сегмента ($L_p = 8$ байт, $L_z(IPv6) = 40$ байт, $L_z(NewIP) = 12$ байт):

$$\eta(IPv6) = 8 / (40 + 8) = 0,167 \text{ (16,7 \%)} \quad (4)$$

$$\eta(NewIP) = 8 / (12 + 8) = 0,400 \text{ (40,0 \%)} \quad (5)$$

$$\Delta\varphi = (0,833 - 0,600) / 0,833 \times 100 \% = 70,0 \% \quad (6)$$

Джиттер (СКО задержки) как мера стабильности передачи:

$$\sigma = \sqrt{[\sum(t_i - \bar{t})^2 / (n - 1)]} \quad (7)$$

где t_i — задержка i -го пакета, $\bar{t}$ — среднее значение задержки, n — число измерений.

Коэффициент доставляемости пакетов:

$$D = (N_{\text{принято}} / N_{\text{отправлено}}) \times 100 \% \quad (8)$$

Таким образом, теоретически обоснованное снижение накладных расходов составляет 70,0 %, а КПД инкапсуляции возрастает в 2,4 раза. Данный аналитический результат будет рассмотрен в контексте экспериментальных данных в разделе 5.

Моделирование проводилось в среде GNS3 версии 3.0.6, развёрнутой на базе виртуальной машины

GNS3 VM под управлением Oracle VirtualBox 7.2 на хост-системе Windows 11. В качестве сетевых узлов использовались Docker-контейнеры на образе python:3.11-alpine.

Топология сети включает четыре узла, соединённых через маршрутизатор R1:

- NodeA (10.0.1.1/24) — генератор исследуемого трафика;
- NodeBG (10.0.2.1/24) — генератор фоновой нагрузки (пакеты 1400 байт $\approx$ Ethernet MTU минус заголовки, непрерывный поток для максимальной нагрузки канала);
- R1 (eth0: 10.0.1.254, eth1: 10.0.2.254, eth2: 10.0.3.254) — маршрутизатор с программной Priority Queuing;
- NodeB (10.0.3.1/24) — приёмник трафика и UDP echo-сервер.

В настоящей работе применена схема UDP echo: NodeB немедленно возвращает каждый полученный пакет обратно отправителю, NodeA измеряет полное время приём-передача с помощью `time.perf_counter()`. Получаемые значения RTT порядка 1–2 мс соответствуют реальным характеристикам виртуализированной сети и являются достоверными. Измерение RTT выполнялось только на этапе 1; на этапах 2,3 фиксировалось время `sendto()` как косвенный показатель загрузки очереди отправителя.

Для корректного сравнения New IP и IPv6 оба протокола маркируются полем DSCP EF через `setsockopt(IP_TOS, 0xb8)` на стороне отправителя. Маршрутизатор R1 реализует Priority Queuing: пакеты портов 9998 (New IP) и 9999 (IPv6) помещаются в единую очередь HIGH, пакеты порта 9997 (фон) — в очередь LOW. Очередь HIGH обслуживается всегда прежде LOW; при переполнении LOW пакеты отбрасываются. Такой подход функционально эквивалентен классу Expedited Forwarding в архитектуре DiffServ и обеспечивает методологически корректное сравнение двух протоколов в идентичных условиях [8].

Эксперимент проводился в три этапа, каждый повторялся трижды для обеспечения статистической достоверности.

Этап 1 — Базовый RTT без QoS. NodeB запускает echo-сервер на портах 9996 (IPv6) и 9995 (New IP). NodeA отправляет 5 серий по 200 пакетов каждого типа напрямую на NodeB, минуя маршрутизатор R1, и измеряет RTT. На этапах 2,3 трафик проходит через R1 и использует порты 9999 (IPv6) и 9998 (New IP), идентифицирующие классы QoS в очередях маршрутизатора. Разделение портов между этапами обусловлено различием сценариев: на этапе 1 порты echo-сервера (9995, 9996) не связаны с логикой QoS и служат исключительно для корректного сопоставления запрос-ответ при измерении

RTT. NodeBG генерирует фоновый поток через R1 на протяжении 320 с. Цель этапа — получить базовую линию задержки в отсутствие приоритизации.

Этап 2 — Равноприоритетная маршрутизация. NodeA направляет трафик через R1. Оба протокола маркируются DSCP EF и попадают в единую очередь HIGH. Фоновый трафик (порт 9997) — в очередь LOW. Фиксируются доставляемость и статистика R1 при фоновой нагрузке ~ 152 Гб за 320 с.

Этап 3 — Критический сигнал при перегрузке канала. Моделируется одновременная передача критической управляющей команды (1000 пакетов, интервал 280 мс — характерный режим для промышленных SCADA-систем с периодом опроса 0,1–1 с) и непрерывного фоновой потока ($\sim 3,62$ Гб за 420 с, два параллельных UDP-потока по 1400 байт). Оба протокола отправляются параллельно в отдельных потоках для обеспечения одновременного присутствия в очереди маршрутизатора. Размер очереди LOW ограничен 5 пакетами для моделирования условий активного вытеснения фоновой нагрузки.

Результаты Этапа 1 представлены в таблице 1. Средний RTT составил 1,342 мс для IPv6 и 1,827 мс для New IP. Высокие значения максимальной задержки (до 1991 мс) и джиттера (27–63 мс) обусловлены влиянием фоновой нагрузки (~ 156 Гб за 320 с) на виртуальные очереди GNS3. На фоне столь высокого джиттера различие средних RTT статистически незначимо: оба протокола ведут себя одинаково без QoS-приоритизации.

Таблица 1.

Этап 1: Реальный RTT без QoS:3 запуска, 1000 пакетов каждого типа

Параметр	Протокол	Запуск 1	Запуск 2	Запуск 3	Среднее
Средний RTT, мс	IPv6	1,406	1,484	1,135	1,342
	New IP	1,118	2,251	2,111	1,827
Макс. RTT, мс	IPv6	974,7	853,4	964,2	930,8
	New IP	988,1	912,4	1991,4	1297,3
Джиттер σ , мс	IPv6	31,730	27,066	30,571	29,79
	New IP	31,324	40,939	63,129	45,13
Доставлено пакетов	IPv6	945/1000	998/1000	995/1000	979/1000
	New IP	995/1000	992/1000	995/1000	994/1000
Фоновый трафик	—	156,1 Гб	156,1 Гб	157,2 Гб	156,5 Гб

Результаты Этапа 2 представлены в таблице 2. Ключевой результат: маршрутизатор R1 во всех трёх запусках пропустил ровно по 1000 пакетов каждого протокола, отбросив при этом ~ 831 000 фоновых пакетов из оче-

реди LOW. Это подтверждает корректность механизма равной приоритизации: оба протокола получали идентичный доступ к каналу.

Таблица 2.

Этап 2: Равноприоритетная маршрутизация DSCP EF:3 запуска, 1000 пакетов

Параметр	Протокол	Запуск 1	Запуск 2	Запуск 3	Среднее
Пропущено R1, пак.	New IP	1000/1000	1000/1000	1000/1000	1000/1000
	IPv6	1000/1000	1000/1000	1000/1000	1000/1000
Доставлено на NodeB, пак.	New IP	1000/1000	782/1000	1000/1000	927/1000
	IPv6	992/1000	758/1000	976/1000	909/1000
Отброшено фона, пак.	—	796 944	815 968	880 960	~831 000
Средняя задержка sendto(), мс	New IP	0,0080	0,0179	0,0039	0,0099
	IPv6	0,0033	0,0038	0,0035	0,0035
Фоновый трафик	—	151,3 ГБ	155,3 ГБ	151,5 ГБ	152,7 ГБ

Доставляемость на NodeB составила 92,7 % для New IP и 90,9 % для IPv6 в среднем по трём запускам. Разброс между запусками является артефактом несинхронности таймеров отправителя и приёмника при последовательной отправке двух потоков: данные R1 подтверждают, что все 1000 пакетов каждого протокола были переданы маршрутизатором, следовательно, потери происходят после R1 из-за рассинхронизации окна приёма на NodeB, а не из-за перегрузки канала. Принципиальный вывод: при одинаковых условиях QoS протоколы New IP и IPv6 неразличимы по доставляемости.

R1 гарантированно пропустил все 1000 пакетов обоих протоколов во всех трёх запусках при активном отбросе фонового трафика (~187 845 пакетов в среднем за сеанс). Потери на NodeB составили 0 % для обоих протоколов во всех запусках. Результаты представлены в таблице 3.

Средние задержки практически идентичны: 0,0566 мс для New IP и 0,0554 мс для IPv6 — разница 2,2 % статистически незначима на фоне джиттера ~0,117 мс. Максимальные задержки также сопоставимы: 1,069 мс для New IP и 0,950 мс для IPv6.

Ключевой результат Этапа 3: механизм Priority Queuing с DSCP EF обеспечивает стопроцентную доставку критических пакетов обоих протоколов даже

Таблица 3.

Этап 3: Критический сигнал при активном вытеснении фона, равный QoS DSCP EF: 3 запуска, 1000 пакетов, интервал 280 мс

Параметр	Протокол	Запуск 1	Запуск 2	Запуск 3	Среднее
Пропущено R1, пак.	New IP	1000/1000	1000/1000	1000/1000	1000/1000
	IPv6	1000/1000	1000/1000	1000/1000	1000/1000
Потери на NodeB, %	New IP	0,0	0,0	0,0	0,0
	IPv6	0,0	0,0	0,0	0,0
Avg задержка sendto(), мс	New IP	0,0560	0,0594	0,0544	0,0566
	IPv6	0,0532	0,0589	0,0540	0,0554
Макс. задержка, мс	New IP	1,2822	0,9548	0,9691	1,069
	IPv6	0,7166	0,9114	1,2206	0,950
Джиттер σ, мс	New IP	0,1163	0,1176	0,1208	0,1182
	IPv6	0,1094	0,1197	0,1175	0,1155
Отброшено фона R1, пак.	—	180 225	189 071	194 240	~187 845
Фоновый трафик	—	3,59 ГБ	3,64 ГБ	3,62 ГБ	3,62 ГБ

при интенсивном фоновом трафике, вытесняя ~187 845 фоновых пакетов за сеанс. Увеличение выборки до 1000 пакетов обеспечивает статистическую достоверность вывода об отсутствии разницы между протоколами: доверительный интервал для доли потерь при $n = 1000$ составляет $\pm 3,1$ % при уровне значимости 0,05, что подтверждает нулевые потери как устойчивый результат. Таким образом, New IP и IPv6 при равных условиях QoS статистически неразличимы по всем измеренным показателям.

Полученные результаты позволяют сформулировать три ключевых вывода.

Первый вывод касается аналитического преимущества New IP в эффективности инкапсуляции. Снижение накладных расходов на 70,0 % является математически доказанным фактом, не зависящим от программной реализации. Для узкополосных каналов LoRaWAN и NB-IoT с полосой 0,3–50 кбит/с это означает двукратное увеличение числа обслуживаемых устройств в ячейке при

Таблица 4.
Сводное сравнение New IP и IPv6

Показатель	IPv6	New IP	Вывод
Размер заголовка IoT, байт	40	12	↓ 70,0 % (аналитика)
КПД инкапсуляции η ($L_p = 8$ байт)	16,7 %	40,0 %	↑ в 2,4 раза (аналитика)
Средний RTT без QoS, мс	1,342	1,827	Статист. неразлично
Доставлено через R1 (Этап 2)	1000/1000	1000/1000	Одинаково
Потери крит. пакетов, NodeB (Этап 3)	0 %	0 %	Одинаково
Защита от вытеснения фоном (R1)	1000/1000	1000/1000	Одинаково при DSCP EF
Отброшено фона R1 (Этап 3)	—	—	~187 845 пакетов

том же объёме передаваемых данных [2, 7]. Данное преимущество New IP является реальным и значимым для практических IoT-приложений в контексте сетей 6G [7, 9].

Второй вывод: при равных условиях QoS сетевое поведение New IP и IPv6 статистически неразлично во всех трёх этапах эксперимента. Это означает, что декларируемые в ряде работ преимущества New IP по задержке и надёжности доставки достигаются исключительно за счёт механизма приоритизации, а не за счёт свойств самого протокола. Тот же результат достигается стандартным DiffServ/DSCP EF применительно к IPv6 [8]. Следовательно, для достижения детерминированной доставки в современных сетях, достаточно корректно настроенного QoS на базе существующих стандартов [6, 8].

Третий вывод: предложенная методика функциональной эмуляции New IP посредством DSCP EF и Priority Queuing на программном маршрутизаторе воспроизводима, не требует специализированного оборудования и пригодна для дальнейших исследований перспективных протоколов Network 2030 [3, 4] в условиях отсутствия их коммерческих реализаций.

Ограничения исследования: среда GNS3/Docker вносит джиттер виртуализации порядка 10–30 мс, не характерный для аппаратных реализаций. Размер заголовка New IP (12 байт) принят как аналитическая оценка на ос-

нове документов ITU-T, реальная реализация может отличаться [3, 4]. Измерение времени `sendto()` на этапах 2–3 является показателем нагрузки очереди ОС, а не сетевой задержки.

Объём фонового трафика на этапах 1–2 и этапе 3 существенно различается. Это обусловлено принципиальным различием задач этапов. На Этапах 1–2 фоновый трафик генерировался без задержек между пакетами с целью создания максимальной нагрузки на виртуальную среду и проверки поведения протоколов в условиях насыщения канала. На этапе 3 задача иная: требуется не максимальная нагрузка, а активное вытеснение фонового трафика механизмом Priority Queuing. При неограниченной скорости генерации Python-маршрутизатор R1 перегружался настолько, что переставал успевать обрабатывать и критические пакеты — они также терялись, что делало сравнение некорректным. Снижение интенсивности фонового потока до уровня ~3,62 Гб за сеанс с сохранением двух параллельных UDP-потоков обеспечивает стабильный отброс ~187 845 фоновых пакетов за счёт ограничения очереди LOW до 5 пакетов, при этом гарантируя полную доставку критических пакетов HIGH-очереди. Таким образом, разница в объёме фона отражает разные исследовательские задачи этапов, а не методологическую непоследовательность.

В работе разработана и апробирована воспроизводимая методика функциональной эмуляции протокола New IP на базе GNS3, Docker и Python 3.11. Проведено трёхкратное повторение трёх серий экспериментов с суммарным фоновым трафиком свыше 900 Гб.

Гипотеза 1 подтверждена аналитически. Заголовок переменной длины New IP обеспечивает снижение накладных расходов на инкапсуляцию на 70,0 % по сравнению с IPv6, увеличивая КПД канала с 16,7 % до 40,0 % при полезной нагрузке 8 байт. Данное преимущество актуально для узкополосных IoT-сетей, таких как LoRaWAN, NB-IoT, Zigbee [2, 7, 9].

Гипотеза 2 подтверждена экспериментально: при равных условиях QoS сетевое поведение New IP и IPv6 статистически неразлично. Маршрутизатор R1 обеспечил стопроцентную пропускную способность обоих протоколов при отбросе ~831 000 фоновых пакетов. При выборке 1000 критических пакетов потери на приёмнике составили 0 % для обоих протоколов во всех трёх запусках при фоновой нагрузке ~3,62 Гб за 420 с и активном отбросе ~187 845 фоновых пакетов маршрутизатором R1.

ЛИТЕРАТУРА

1. Deering S., Hinden R. Internet Protocol, Version 6 (IPv6) Specification // RFC 8200. IETF, 2017. URL: <https://www.rfc-editor.org/rfc/rfc8200> (дата обращения: 05.03.2026).
2. Тимохина А.А., Куваева Е.Н. Протокол IPv6: обзор ключевых характеристик // Вестник науки. 2025. № 8 (89). Т. 2. С. 338–343. ISSN 2712-8849. (дата обращения: 05.03.2026).
3. ITU-T Focus Group on Technologies for Network 2030. A Blueprint of Technology, Applications and Market Drivers Towards the Year 2030 and Beyond. Geneva: ITU, 2019. URL: https://www.itu.int/en/ITU-T/focusgroups/net2030/Documents/White_Paper.pdf (дата обращения: 11.03.2026).
4. ICANN Office of the CTO. An Analysis of the «New IP» Proposal to the ITU-T. OCTO-017. October 2020. URL: <https://www.icann.org/en/system/files/files/octo-017-27oct20-en.pdf> (дата обращения: 11.03.2026).
5. Internet Society. Discussion Paper: An Analysis of the «New IP» Proposal to the ITU-T. 2020. URL: <https://www.internetsociety.org/resources/doc/2020/discussion-paper-an-analysis-of-the-new-ip-proposal-to-the-itu-t/> (дата обращения: 19.03.2026).
6. Davie B., Charny A., Bennet J.C.R. et al. An Expedited Forwarding PHB (Per-Hop Behavior) // RFC 3246. IETF, 2002. URL: <https://www.rfc-editor.org/rfc/rfc3246> (дата обращения: 19.03.2026).
7. Тихвинский В., Девяткин Е., Белявский В., Смирнов Ю. Архитектура сетей 6G: принципы и особенности построения. Ч. 1 // Первая миля. 2022. № 3. DOI: 10.22184/2070-8963.2022.103.3.50.56 (дата обращения: 23.03.2026).
8. Blake S., Black D., Carlson M. et al. An Architecture for Differentiated Services // RFC 2475. IETF, 1998. URL: <https://www.rfc-editor.org/rfc/rfc2475> (дата обращения: 23.03.2026).
9. Росляков А.В. Будущие сети: обзор подходов к новой телекоммуникационной парадигме // Электросвязь. 2020. № 9. С. 30–37. (дата обращения: 26.03.2026).

© Адаев Николай Олегович (adaev.kol@yandex.ru); Пожидаев Николай Андреевич (pozhideev63nik@gmail.com);
Маркарян Петрос Габриелович (markaryan.petros@yandex.ru); Таранец Максим Евгеньевич (mtaranetz@yandex.ru);
Малахов Сергей Валерьевич (malakhov.sv@ssau.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

СТРУКТУРНЫЕ МОДЕЛИ LLM-АГЕНТОВ И ИХ ОГРАНИЧЕНИЯ С ТОЧКИ ЗРЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ: ОБЗОР И СРАВНИТЕЛЬНЫЙ АНАЛИЗ

STRUCTURAL MODELS OF LLM AGENTS AND THEIR SECURITY LIMITATIONS: A REVIEW AND COMPARATIVE ANALYSIS

Alkattan Sameh

Summary. The paper reviews structural models of LLM agents and analyzes their limitations from the information-security perspective. The study focuses on agent components such as memory, planner, action space, external tools, data sources and runtime environment. The methodology is based on a structured review of selected publications, specifications and Russian-language sources published in 2022–2026. The analysis uses criteria related to trust boundaries, source authorization, data isolation, tool-call control, policy enforcement and auditability. The main result is a comparative matrix of structural models and a classification of the reviewed sources by their level of security coverage. The paper shows that many structural models describe how agents operate, but do not sufficiently formalize how agents should be protected against untrusted inputs and unauthorized actions.

Keywords: LLM agent, structural model, information security, prompt injection, access control, trust boundary, security policy, tool calls, agentic systems.

Алькаттан Самех

Аспирант, Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Санкт-Петербургский государственный
университет телекоммуникаций
имени профессора М.А. Бонч-Бруевича»
sameh.kt94@gmail.com

Аннотация. Целью статьи является обобщение современных структурных моделей LLM-агентов и выявление их ограничений с точки зрения информационной безопасности. В работе рассматриваются модели, описывающие агент как систему с памятью, планировщиком, пространством действий, внешними инструментами, источниками данных и средой выполнения. Методология исследования основана на структурированном обзоре отобранных публикаций 2022–2026 годов, спецификаций и русскоязычных работ по архитектуре и кибербезопасности агентных систем. Для анализа применяются критерии наличия границ доверия, авторизации источников, изоляции данных, контроля вызовов инструментов, принудительного применения политик и аудируемости действий. Результатом является сравнительная матрица моделей, классификация источников по уровню учёта безопасности и описание исследовательского разрыва: значительная часть архитектурных моделей хорошо описывает функциональные элементы агента, но не содержит достаточной формализации защитных механизмов. Показано, что перспективным направлением дальнейших исследований является переход от описания архитектуры агента к модели контроля цепочки «источник данных — интерпретация — действие», в которой безопасность оценивается не только по факту обнаружения вредоносного текста, но и по способности предотвратить несанкционированное привилегированное действие.

Ключевые слова: LLM-агент, структурная модель, информационная безопасность, prompt-инъекция, контроль доступа, доверенная граница, политика безопасности, инструментальные вызовы, агентные системы.

Введение

Переход больших языковых моделей от режима диалогового помощника к режиму агентной системы изменил содержание задачи информационной безопасности. Если в обычном диалоге результатом ошибки модели является некорректный ответ, то в агентной системе ошибка интерпретации может привести к вызову внешнего инструмента, передаче данных, изменению состояния внешнего сервиса или запуску программного действия. Поэтому структурная модель LLM-агента становится не только средством описания

функциональности, но и основой для определения границ доверия, источников воздействия и точек контроля.

В современных работах под агентом на базе большой языковой модели обычно понимается система, которая сочетает языковое рассуждение, память, планирование, обращение к внешним источникам данных и выполнение действий через инструменты. Такой подход отражён в модели CoALA, где языковой агент описывается через компоненты памяти, пространство действий и процесс принятия решений [12]. Близкую идею связывания рассуждения и действия показывает ReAct: модель череду-

ет рассуждения с обращением к внешним источникам или средам [14]. Эти модели полезны для объяснения архитектуры агента, но сами по себе не решают вопрос о том, каким образом недоверенные данные должны быть отделены от управляющих инструкций.

Практические агентные экосистемы усилили данную проблему. Протокол MCP описывает архитектуру подключения LLM-приложений к внешним инструментам и источникам данных; он задаёт элементы *host*, *client*, *server*, *tools* и *resources* [9]. Такая архитектура удобна для интеграции, но одновременно расширяет поверхность атаки, поскольку данные из внешней среды могут влиять на выбор инструментов и параметры вызовов. Русскоязычные публикации 2025–2026 годов также фиксируют рост интереса к архитектуре LLM-агентов, их кибербезопасности, уязвимостям MCP-экосистемы, методам атак и моделированию угроз [1; 2; 3; 4; 5].

Цель настоящей статьи состоит в том, чтобы сопоставить существующие структурные модели LLM-агентов и показать, в какой степени они учитывают информационную безопасность. В отличие от общих обзоров, работа сосредоточена не на перечне всех известных агентов, а на сравнении структурных подходов по признакам, важным для защиты: наличие границ доверия, авторизации источников, изоляции данных, контроля действий, принудительного применения политик и аудируемости. Такая постановка позволяет сформировать основу для последующих работ, связанных с *source-to-action firewall* для высокопривилегированных агентов.

Методика отбора и критерии анализа

Исследование выполнено как структурированный обзор. В выборку были включены источники, опубликованные в 2022–2026 годах, а также одна официальная спецификация протокола. Приоритет отдавался работам, в которых агентная система описывается через архитектурные элементы, инструментальные вызовы, память, протоколы подключения, бенчмарки безопасности или формальные механизмы защиты. Отдельно учитывались русскоязычные публикации, поскольку они важны для последующего оформления диссертационного исследования и обоснования актуальности темы в российском научном поле.

Отбор источников был ограничен задачей предварительного сопоставления моделей, а не претендовал на полный систематический обзор всей литературы. Поэтому далее под статистикой понимается не глобальная доля публикаций в мировой литературе, а распределение внутри контролируемой аналитической выборки. Такой подход позволяет избежать неподтверждённых обобщений и одновременно показать, какие типы моделей чаще используются как структурная основа, а какие уже включают элементы безопасности.

Для оценки источников применялись шесть критериев. Первый критерий — наличие явной структурной декомпозиции агента. Второй — фиксация внешних инструментов и каналов данных как отдельных элементов модели. Третий — наличие границ доверия между пользователем, внешними источниками, моделью, инструментами и средой выполнения. Четвёртый — возможность описывать привилегированные действия. Пятый — наличие механизмов политики или принудительного контроля. Шестой — возможность использовать модель для количественной оценки безопасности.

Таблица 1. Критерии анализа структурных моделей LLM-агентов

Критерий	Содержание критерия	Значение для информационной безопасности
Структурная декомпозиция	Выделяются память, планировщик, действия, инструменты или среда	Позволяет определить объект защиты и точки контроля
Инструменты и внешние данные	Модель учитывает каналы обращения к сервисам, документам, веб-страницам или API	Позволяет анализировать <i>indirect prompt injection</i> и <i>tool abuse</i>
Границы доверия	Разделяются доверенные и недоверенные источники инструкций и данных	Необходимо для <i>source authorization</i> и <i>data isolation</i>
Привилегированные действия	Модель различает действия с внешним эффектом и низкорисковые действия	Позволяет измерять несанкционированные действия
Политики и принуждение	Предусмотрена проверка правил до выполнения действия	Создаёт основу для <i>firewall</i> / <i>reference monitor</i>
Измеримость	Модель поддерживает показатели безопасности и полезности	Позволяет оценивать эффективность защитных мер

Основные структурные модели LLM-агентов

CoALA является удобной отправной точкой для обзора, поскольку она связывает современные языковые агенты с традицией когнитивных архитектур. В рамках CoALA агент описывается как система, имеющая память, пространство действий и процедуру выбора действий [12]. Для задачи информационной безопасности эта модель важна тем, что позволяет указать, где в агенте находятся память, внутренние действия и внешние действия. Однако CoALA не формулирует собственную модель угроз и не задаёт механизмов контроля источников, политик доступа или изоляции данных. Следовательно, при использовании CoALA в диссертационном исследовании её целесообразно рассматривать как базовую

вый структурный слой, который должен быть дополнен средствами безопасности.

ReAct описывает агентное поведение как чередование рассуждений и действий [14]. В практическом смысле это близко к поведению многих инструментальных агентов: модель рассуждает о задаче, выбирает действие, получает результат из среды и продолжает рассуждение. Преимущество ReAct состоит в простоте и интерпретируемости траектории; ограничение — в отсутствии встроенной проверки того, является ли источник результата доверенным и имеет ли он право задавать агенту новые инструкции. Поэтому ReAct полезен для экспериментов и демонстрации source-to-action цепочки, но требует внешнего слоя безопасности.

AutoGen дополняет эту картину многоагентной архитектурой: отдельные агенты могут вести диалог, использовать инструменты, подключать человека и работать в разных режимах взаимодействия [13]. Для задач безопасности это означает появление не только одного планировщика, но и цепочки межагентных сообщений, где источник инструкции и ответственность за действие могут быть распределены между несколькими агентами. Поэтому многоагентные фреймворки требуют дополнительных критериев аудита, авторизации источников и контроля передачи данных между агентами.

MCP представляет другой тип структурной модели. В отличие от CoALA и ReAct, он не описывает когнитивную организацию агента, а задаёт протокольную архитектуру подключения LLM-приложений к инструментам и источникам данных [9]. С точки зрения безопасности важны элементы host, client, server, resources и tools, а также возможность авторизации на транспортном уровне. При этом сама спецификация не превращает систему в защищённую автоматически: безопасность зависит от доверия к серверам, описаниям инструментов, политик доступа, проверки аргументов и контроля результатов. Поэтому MCP следует рассматривать как

важную практическую архитектуру, которая выявляет реальные границы доверия.

Fides, в отличие от предыдущих моделей, является уже непосредственно защитно-ориентированным подходом. В нём используется контроль информационных потоков, метки конфиденциальности и целостности, а также детерминированное применение политик для планировщика агента [7]. Это существенно меняет характер модели: агент рассматривается не только как функциональный исполнитель, но и как объект формального контроля. Аналогичное направление развивает PCAS, где состояние агентной системы представляется в виде графа зависимостей, а правила политики применяются через reference monitor до исполнения действия [10]. Эти подходы демонстрируют, что безопасность агента должна контролироваться не только через prompt, но и через формальную связь между источником, контекстом и действием.

Отдельное место занимают бенчмарки и guardrail-системы. AgentDojo предоставляет среду для оценки prompt-injection атак и защит в агентных задачах с внешними инструментами [8]. InjecAgent показывает, что косвенные prompt-инъекции могут реализовываться через инструменты и внешние данные [15]. LlamaFirewall представляет firewall-подобную guardrail-систему для AI-агентов и включает компоненты защиты от prompt injection, misalignment и небезопасного кода [6]. Эти источники не всегда являются структурными моделями в узком смысле, но важны для понимания того, какие элементы архитектуры должны быть контролируемы в экспериментальной постановке.

Статистика по контролируемой выборке источников

Для первичного количественного анализа была сформирована выборка из пятнадцати источников. В неё вошли общие структурные модели, протокольные



Рис. 1. Логика обзора структурных моделей и ограничений безопасности

Таблица 2.

Сравнительная матрица структурных моделей и подходов

Модель / подход	Основная архитектура	Инструменты и среда	Учет безопасности	Ограничение
CoALA	Память + пространство действий + принятие решений	Да, через внутренние и внешние действия	Низкий	Нет явной модели угроз и политик
ReAct	Чередование рассуждения и действия	Да, через вызовы внешних источников	Низкий	Нет проверки полномочий источника
MCP	Host — client — server — tools/resources	Да, протокольный уровень	Средний	Безопасность зависит от реализации политик
AgentDojo	Агент выполняет задачи с инструментами над недоверенными данными	Да, в бенчмарке	Средний/высокий	Среда оценки, а не формальная модель защиты
Fides / IFC	Планировщик + метки конфиденциальности и целостности	Да	Высокий	Требует корректной разметки и политики
PCAS	Граф зависимостей + язык политик + reference monitor	Да	Высокий	Требует инструментирования выполнения
LlamaFirewall	Сканеры и guardrails для агентного выполнения	Да	Высокий	Нужна проверка устойчивости к адаптивным атакам

архитектуры, бенчмарки, формальные модели защиты, firewall-подобные решения и русскоязычные работы по LLM-агентам. Источники классифицировались не по количеству цитирований и не по рыночной популярности, а по тому, какую роль они могут играть в будущей диссертационной работе: базовая модель агента, описание протокола, бенчмарк, формальная модель защиты или обзор угроз.

Полученное распределение показывает, что общие архитектурные модели и протокольные описания необходимы для определения объекта исследования, но лишь часть источников содержит полноценные механизмы безопасности. В частности, CoALA и ReAct дают структуру и логику действий, однако не определяют правила авторизации источников. MCP вводит реальные границы между host, client, server и tools, но не является самодостаточным механизмом защиты. Наиболее силь-

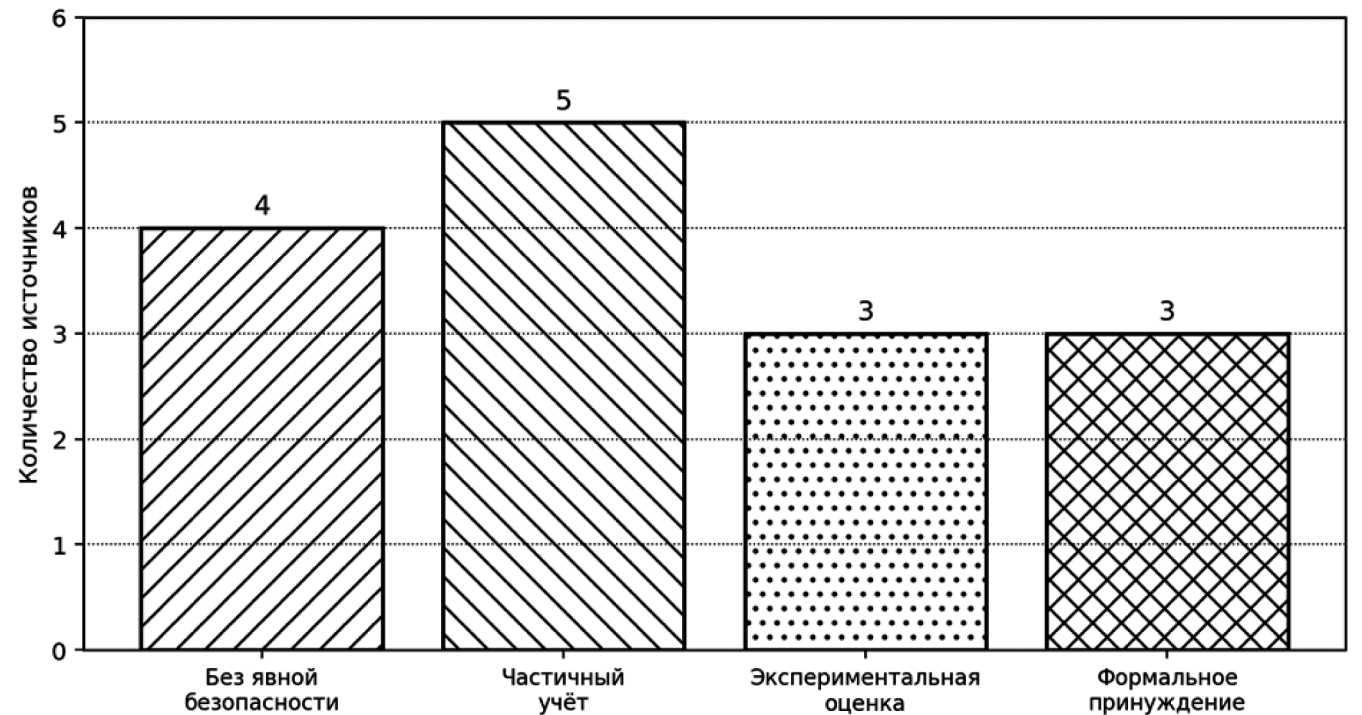


Рис. 2. Распределение источников по уровню учёта безопасности в контролируемой выборке

ный формальный слой представлен подходами Fides, PCAS и контекстной формализацией безопасности LLM-агентов [7; 10; 11].

Распределение по уровню учёта безопасности отражено на рис. 2. Категория «без явной безопасности» включает источники, где агент описывается функционально. Категория «частичный учёт» включает работы и спецификации, в которых безопасность упоминается через авторизацию, угрозы или уязвимости, но не формируется строгий механизм принуждения. Категория «экспериментальная оценка» включает бенчмарки и guardrail-системы. Категория «формальное принуждение» включает модели, в которых действие проверяется по правилам, меткам или графу зависимостей до исполнения.

Таблица 3.

Распределение источников по роли в исследовании

Категория	Примеры источников	Количество	Доля в выборке
Базовые структурные модели	CoALA, ReAct, AutoGen, русскоязычная архитектура LLM-агентов	4	26,7 %
Протокольные и инструментальные архитектуры	MCP, работы об MCP-экосистеме	2	13,3 %
Бенчмарки и экспериментальные среды	AgentDojo, InjecAgent	2	13,3 %
Формальные модели и policy enforcement	Fides, PCAS, контекстная модель безопасности	3	20,0 %
Firewall / guardrails и русскоязычные обзоры угроз	LlamaFirewall, работы о кибербезопасности и атаках LLM-агентов	4	26,7 %

Выявленные ограничения с точки зрения информационной безопасности

Первое ограничение состоит в том, что значительная часть моделей описывает агента как функциональную систему, но не задаёт явных доверенных и недоверенных зон. Для задач информационной безопасности этого недостаточно. Если внешний документ, письмо или результат инструмента попадает в контекст планировщика без маркировки источника, агент может интерпретировать недоверенные данные как управляющую инструкцию. В результате функциональная архитектура превращается в канал косвенной prompt-инъекции.

Второе ограничение связано с отсутствием явного различия действий по уровню привилегий. Струк-

турные модели часто говорят о пространстве действий, но не делят действия на низкорисковые и высокорисковые. Между тем для защищаемых систем принципиально важно различать, например, вычисление выражения, чтение документа, отправку сообщения, запись в базу данных или выполнение команды. Без такой классификации невозможно корректно измерить долю несанкционированных привилегированных действий.

Третье ограничение проявляется в слабой связи между моделью агента и измеримыми показателями безопасности. Бенчмарки типа AgentDojo и InjecAgent позволяют измерять успешность атак и полезность агента [8; 15], но не каждая структурная модель автоматически переводится в систему метрик. Для диссертационного исследования важно связать архитектурный слой с показателями: доля успешных атак, доля нарушений политики, доля несанкционированных действий, полезность при атаке и уровень ложных срабатываний.

Четвёртое ограничение связано с тем, что многие защитные меры фокусируются на тексте запроса или результата, тогда как реальная угроза появляется в цепочке от источника к действию. Например, опасным является не сам факт наличия вредоносной фразы в письме, а ситуация, когда это письмо заставляет агента вызвать инструмент отправки данных или чтения конфиденциального профиля. Поэтому future work должен развивать не только prompt-фильтры, но и контроль source-to-action: источник, интерпретация, планировщик, вызов инструмента, внешний эффект.

Пятое ограничение касается многоязычности и способов представления вредоносных инструкций. Существующие guardrails и firewall-подобные решения могут учитывать часть таких атак, но в литературе остаётся открытым вопрос о единой методике оценки устойчивости к русскоязычным, англоязычным, арабоязычным, смешанным и кодированным инструкциям. Это особенно важно для LLM-агентов, работающих в международных организациях или в средах, где документы и переписка представлены на нескольких языках.

Обсуждение: от структурной модели к модели контроля

Проведённый обзор показывает, что наиболее перспективной является не замена существующих моделей, а их объединение в многоуровневую схему. На первом уровне CoALA и близкие архитектурные модели используются для описания агента как системы с памятью, действиями и принятием решений. На втором уровне MCP и другие инструментальные архитектуры позволяют описать внешние каналы, через которые агент получает данные и вызывает инструменты. На третьем уровне Fides, PCAS и контекстные модели безопасности задают

свойства, которым должны соответствовать действия агента.

Для будущей диссертационной работы это даёт естественную логику научных результатов. Первый результат может быть связан со структурной моделью высокопривилегированного агента, в которой явно выделяются доверенные границы, источники данных и привилегированные действия. Второй результат может быть связан с системой показателей и математической моделью оценки безопасности. Третий результат может состоять в методе защиты, реализующем firewall-подобный контроль перед выполнением действия. Такое направление соответствует современному состоянию исследований: оно не ограничивается обнаружением «плохого текста», а контролирует возможность перехода от недоверенного источника к несанкционированному действию.

Практически данная схема может быть реализована через два контрольных барьера. Первый барьер анализирует входные данные, например письмо, документ, веб-страницу, результат инструмента или описание инструмента. Второй барьер проверяет действие, которое агент собирается выполнить. Если инструкция пришла из недоверенного источника и приводит к высокорисковому действию, например отправке данных внешнему адресату, такое действие должно быть заблокировано, ограничено или вынесено на подтверждение пользователя. Подобный подход согласуется с идеей принудительного применения политик, а не только с идеей классификации текста.

Следовательно, основной исследовательский разрыв формулируется следующим образом: существующие структурные модели в значительной степени пригодны для описания агента, но не всегда пригодны для прямой оценки защищённости; существующие защитные модели дают сильные механизмы контроля, но пока не образуют единой прикладной методики для высокопривилегированных LLM-агентов, работающих с внешними данными и инструментами. Заполнение

этого разрыва может стать основой последующей статьи с экспериментом.

Заключение

В статье выполнен обзор и сравнительный анализ структурных моделей LLM-агентов с позиции информационной безопасности. Показано, что базовые модели, такие как CoALA и ReAct, позволяют описать функциональную организацию агента, но не содержат достаточного набора защитных механизмов. Протокольные архитектуры типа MCP вводят реальные границы между приложением, клиентом, сервером и инструментами, однако требуют внешних политик доступа и контроля действий. Наиболее близкими к задаче защиты являются Fides, PCAS, контекстные модели безопасности и firewall-подобные guardrail-системы, поскольку они рассматривают действия агента как объект контроля.

Сравнительный анализ подтверждает, что актуальным направлением является разработка source-to-action модели контроля безопасности для высокопривилегированных LLM-агентов. Такая модель должна учитывать источник инструкции, контекст задачи, доверенность данных, уровень привилегий инструмента и внешний эффект действия. Для первой публикации полученный обзор может использоваться как теоретическое обоснование темы; для следующего этапа целесообразно выполнить эксперимент с email-агентом или инструментальным агентом, сравнив обычный режим работы, текстовый фильтр и action-firewall по показателям доли успешных атак, доли несанкционированных действий, полезности и ложных срабатываний.

Таким образом, основная научная перспектива заключается в переходе от описательных структурных моделей к измеримой и проверяемой модели информационной безопасности LLM-агентов. Именно эта связь — структура агента, показатели безопасности и механизм принудительного контроля действий — может стать основой дальнейшего диссертационного исследования.

ЛИТЕРАТУРА

1. Евграфов В.А., Нутфуллин Б.М., Намиот Д.Е. Методы атак и защиты в агентных системах на основе больших языковых моделей // International Journal of Open Information Technologies. 2026. Т. 14. № 5. С. 1–8. URL: <https://cyberleninka.ru/article/n/metody-atak-i-zaschity-v-agentnyh-sistemah-na-osnove-bolshih-yazykovyh-modeley> (дата обращения: 16.05.2026).
2. Лапоница О.Р., Костин Р.Н. Разработка программного обеспечения моделирования угроз для систем на базе LLM-агентов // International Journal of Open Information Technologies. 2025. URL: <https://cyberleninka.ru/article/n/razrabotka-programmnogo-obespecheniya-modelirovaniya-ugroz-dlya-sistem-na-baze-llm-agentov> (дата обращения: 16.05.2026).
3. Намиот Д.Е., Ильюшин Е.А. Архитектура LLM агентов // International Journal of Open Information Technologies. 2025. URL: <https://cyberleninka.ru/article/n/arhitektura-llm-agentov> (дата обращения: 16.05.2026).
4. Намиот Д.Е., Ильюшин Е.А. О кибербезопасности ИИ-агентов // International Journal of Open Information Technologies. 2025. URL: <https://cyberleninka.ru/article/n/o-kiberbezopasnosti-ii-agentov> (дата обращения: 16.05.2026).
5. Намиот Д.Е., Ильюшин Е.А. Уязвимости экосистемы MCP // International Journal of Open Information Technologies. 2025. URL: <https://cyberleninka.ru/article/n/uyazvimosti-ekosistemy-mcp> (дата обращения: 16.05.2026).

6. Chennabasappa S., Nikolaidis C., Song D. et al. LlamaFirewall: An Open Source Guardrail System for Building Secure AI Agents. 2025. arXiv:2505.03574. URL: <https://arxiv.org/abs/2505.03574> (accessed: 16.05.2026).
7. Costa M., Köpf B., Kolluri A. et al. Securing AI Agents with Information-Flow Control. 2025. arXiv:2505.23643. URL: <https://arxiv.org/abs/2505.23643> (accessed: 16.05.2026).
8. DeBenedetti E., Zhang J., Balunović M. et al. AgentDojo: A Dynamic Environment to Evaluate Prompt Injection Attacks and Defenses for LLM Agents // Advances in Neural Information Processing Systems. 2024. URL: <https://arxiv.org/abs/2406.13352> (accessed: 16.05.2026).
9. Model Context Protocol. Specification. 2025. URL: <https://modelcontextprotocol.io/specification/2025-06-18> (accessed: 16.05.2026).
10. Palumbo N., Choudhary S., Choi J. et al. Policy Compiler for Secure Agentic Systems. 2026. arXiv:2602.16708. URL: <https://arxiv.org/abs/2602.16708> (accessed: 16.05.2026).
11. Siu V., He J., Montgomery K. et al. A Framework for Formalizing LLM Agent Security. 2026. arXiv:2603.19469. URL: <https://arxiv.org/abs/2603.19469> (accessed: 16.05.2026).
12. Sumers T.R., Yao S., Narasimhan K., Griffiths T.L. Cognitive Architectures for Language Agents // Transactions on Machine Learning Research. 2024. URL: <https://openreview.net/forum?id=1i6ZCvflQJ> (accessed: 16.05.2026).
13. Wu Q., Bansal G., Zhang J. et al. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation Framework. 2023. arXiv:2308.08155. URL: <https://arxiv.org/abs/2308.08155> (accessed: 16.05.2026).
14. Yao S., Zhao J., Yu D. et al. ReAct: Synergizing Reasoning and Acting in Language Models // Proceedings of ICLR 2023. URL: <https://arxiv.org/abs/2210.03629> (accessed: 16.05.2026).
15. Zhan Q., Liang Z., Ying Z., Kang D. InjecAgent: Benchmarking Indirect Prompt Injections in Tool-Integrated Large Language Model Agents // Findings of the Association for Computational Linguistics: ACL 2024. P. 10471–10506. URL: <https://aclanthology.org/2024.findings-acl.624/> (accessed: 16.05.2026).

© Алькаттан Самех (sameh.kt94@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»

ЭФФЕКТИВНАЯ ОБРАБОТКА МАЛЫХ ДАННЫХ В ТРЁХМЕРНОЙ ОБЪЁМНОЙ СЕГМЕНТАЦИИ НЕЙРОИЗОБРАЖЕНИЙ: ОГРАНИЧЕНИЕ ПАРАМЕТРОВ КАК СТРУКТУРНЫЙ РЕГУЛЯРИЗАТОР

EFFICIENT SMALL DATA PROCESSING IN THREE-DIMENSIONAL VOLUMETRIC SEGMENTATION OF NEUROIMAGES: PARAMETER CONSTRAINT AS A STRUCTURAL REGULATOR

Apana Kenneth Ayinbuno
Yu. Shichkina

Summary. Modern three-dimensional (3D) convolutional neural networks (CNNs) have achieved remarkable success in volumetric medical image segmentation, but such highly parameterized architectures typically require massive annotated datasets ($N > 100$) to converge without catastrophic overfitting. In highly specialized clinical domains, such as the spatial localization of specific cortical areas for neuromodulation, collecting such datasets is virtually impossible, leading to the extreme «small data» regime ($N \approx 10$). This paper proposes a data-efficient 3D segmentation framework that deliberately limits architectural capacity to ensure structural regularization. By mathematically reducing the number of base filters in a 3D U-Net, we achieve a 16-fold reduction in parametric complexity (~ 1.4 million parameters). When combined with a stochastic class-balanced patch sampling algorithm ($p \geq 0.6$) to address severe volumetric class imbalance, the proposed model is successfully trained from scratch without knowledge transfer. Evaluated on a pilot dataset of $N = 10$ structural MRI images using a strict leave-one-out (LOSO) cross-validation protocol, the lightweight model achieved an average Dice similarity coefficient (DSC) of 0.838 with a peak value of 0.969. The results demonstrate that strict parameter pruning is a highly effective regularization strategy for extracting complex 3D topological features under severely data-poor conditions.

Keywords: deep learning, 3D U-Net, small data, volumetric segmentation, Vapnik–Chervonenkis dimension, structural regularization.

Апана Кеннет Айнбуно

Аспирант, Санкт-Петербургский государственный
электротехнический университет «ЛЭТИ»
kenneth.89@mail.ru

Шичкина Юлия Александровна

Доктор технических наук, профессор,
Санкт-Петербургский государственный
электротехнический университет «ЛЭТИ»
strange.y@mail.ru

Аннотация. Современные трёхмерные (3D) свёрточные нейронные сети (CNN) достигли выдающихся успехов в сегментации объёмных медицинских изображений, но столь сильно параметризованные архитектуры, как правило, требуют массивных аннотированных наборов данных ($N > 100$) для сходимости без катастрофического переобучения. В узкоспециализированных клинических областях, таких как пространственная локализация конкретных корковых зон для нейромодуляции, сбор подобных наборов данных практически невозможен, что приводит к крайнему режиму «малых данных» ($N \approx 10$). В данной работе предлагается эффективная по данным среда трёхмерной сегментации, которая целенаправленно ограничивает архитектурную ёмкость для обеспечения структурной регуляризации. Путём математического сокращения числа базовых фильтров трёхмерной U-Net мы добиваемся 16-кратного снижения параметрической сложности ($\sim 1,4$ млн параметров). В сочетании со стохастическим алгоритмом сбалансированной по классам выборки патчей ($p \geq 0,6$) для устранения серьёзного объёмного дисбаланса классов предлагаемая модель успешно обучается с нуля без использования переноса знаний. Оценённая на пилотном наборе из $N = 10$ структурных МРТ-снимков с применением строгого протокола кросс-валидации с исключением одного субъекта (LOSO), облегчённая модель достигла среднего коэффициента сходства Дайса (DSC) 0,838 при пиковом значении 0,969. Результаты показывают, что строгое сокращение параметров является высокоэффективной стратегией регуляризации для извлечения сложных трёхмерных топологических признаков в условиях острой нехватки данных.

Ключевые слова: глубокое обучение, трёхмерная U-Net, малые данные, объёмная сегментация, размерность Вепника–Червоненкиса, структурная регуляризация.

Автоматическая семантическая сегментация трёхмерной (3D) медицинской визуализации представляет собой классическую задачу компьютерного зрения. С появлением полносвёрточных сетей (FCN) архитектуры, такие как 3D U-Net [1] и V-Net [2], стали стандартом де-факто для извлечения объёмных топо-

логических признаков. Современные передовые (SOTA) платформы, в частности nnU-Net [3] и визуальные трансформеры (UNETR) [4], демонстрируют исключительную точность на общедоступных эталонных наборах, но фундаментальное вычислительное узкое место ограничивает повсеместное клиническое внедрение этих моделей

и это проблема малых данных. Стандартные передовые трёхмерные архитектуры параметризованы десятками миллионов весов. Обучение столь высокорежимных моделей требует массивных аннотированных наборов данных ($N > 100-1000$) для адекватного ограничения математического пространства решений.

В специализированных медицинских приложениях — таких как выделение дорсолатеральной префронтальной коры (DLPFC) для навигации транскраниальной магнитной стимуляции (ТМС) — разметка одного трёхмерного объёма требует часов работы эксперта-нейроанатома. Вследствие этого исследователи зачастую ограничены пилотными когортами ($N \leq 10$). При применении в этом крайнем режиме нехватки данных стандартные сильно параметризованные модели страдают катастрофическим переобучением, запоминая обучающий шум и показывая почти нулевую обобщающую способность на новых данных.

Данная работа посвящена абсолютной нижней границе достаточности данных в трёхмерной медицинской сегментации. Мы выдвигаем гипотезу, что за счёт решительного переноса фокуса оптимизации с «максимальной репрезентативной ёмкости» на «эффективность по данным» структурно ограниченная трёхмерная CNN способна надёжно выучивать сложные кортикальные признаки с нуля на минимальном наборе данных.

Границы обобщения и сложность модели

Необходимость архитектурного сокращения может быть формализована с помощью теории статистического обучения. В рамках границ Вапника–Червоненкиса (VC) [5] ошибка обобщения E_{gen} (ожидаемое качество на невиданных данных) теоретически ограничена эмпирической ошибкой обучения E_{train} плюс штрафным членом сложности:

$$E_{gen} \leq E_{train} + \sqrt{\frac{h \left(\log \left(\frac{2N}{h} \right) + 1 \right) - \log \left(\frac{\eta}{4} \right)}{N}} \quad (1)$$

где N — число обучающих примеров, h — размерность Вапника–Червоненкиса (сильно коррелирующая с количеством обучаемых параметров W), а $1-\eta$ — статистическая достоверная вероятность.

В режиме $N \approx 10$, если ёмкость модели $h \rightarrow$ «велика» (например, стандартный трёхмерный остов на основе VGG16, содержащий $\approx 33 \times 10^6$ параметров), штраф за сложность возрастает экспоненциально. Математическое неравенство вырождается в тривиальное утверждение ($E_{gen} \leq E_{train} + \infty$), подразумевая отсутствие теоретической гарантии алгоритмического обобщения (т.е. тяжёлое переобучение). Для активной минимизации E_{gen}

алгоритм должен одновременно минимизировать E_{train} посредством градиентного спуска и строго минимизировать h .

Масштабирование фильтров и сокращение параметров

Пусть W_l — полное количество обучаемых параметров в конкретном трёхмерном свёрточном слое l :

$$W_l = F_{in} \times F_{out} \times k_x \times k_y \times k_z \quad (2)$$

где F — число каналов признаков, а k — пространственные размеры объёмного ядра ($3 \times 3 \times 3$).

В стандартных архитектурах трёхмерных сетей базовое количество фильтров обычно инициализируется как $F_{base} = 64$. В нашей предложенной облегчённой архитектуре базовое количество фильтров строго ограничено значением $F_{base} = 16$. Поскольку число параметров квадратично зависит от количества фильтров в промежуточных скрытых слоях (где отображение, как правило, имеет вид $F \rightarrow 2F$), общее снижение структурной сложности R моделируется как:

$$R \propto \left(\frac{F_{standard}}{F_{light}} \right)^2 = \left(\frac{64}{16} \right)^2 = 16$$

Таблица 1.

Теоретическое и вычислительное сравнение стандартных передовых (SOTA) объёмных архитектур и предложенной структурно ограниченной модели

Тип архитектуры	Базовое число фильтров	Макс. число фильтров (узкое место)	Всего параметров	Требуемый пиковый объём видеопамяти
Стандартная 3D U-Net / V-Net	64	512	$\approx 22\,000\,000$	> 16 Гб
Предложенная облегчённая модель	16	128	$\approx 1\,400\,000$	4,3 Гб

Источник: составлено авторами

Это теоретическое 16-кратное сокращение существенно ужесточает уравнение VC-границы. Оно действует как мощный структурный регуляризатор, математически вынуждая модель выучивать лишь наиболее значимые, инвариантные топологические признаки (например, сильные градиентные перепады на границах серого и белого вещества) вместо запоминания высокочастотных артефактов сканера, специфичных для минимального обучающего набора.

Облегчённая архитектура трёхмерной U-Net

Предлагаемый конвейер инференса глубокого обучения состоит из симметричной структуры кодер-декодер, содержащей приблизительно $1,4 \times 10^6$ обучаемых параметров.

Кодер: состоит из четырёх блоков свёрток $3 \times 3 \times 3$, за каждым из которых следует субдискретизация максимума (англ. Max Pooling) $2 \times 2 \times 2$. Глубина карт признаков прогрессивно нарастает: $16 \rightarrow 32 \rightarrow 64 \rightarrow 128$.

Декодер: использует трёхмерные транспонированные свёртки для повышения пространственного разрешения карт признаков. Пропускные связи из кодера конкатенируются с соответствующими слоями декодера, чтобы сохранить высокочастотные пространственные детали, критически важные для точного определения границ.

Оптимизация: модель оптимизируется с помощью составного функционала потерь $J(\theta) = L_{BCE} + (1 - DSC)$, который уравнивает устойчивость попиксельной бинарной кросс-энтропии с максимизацией глобально пространственного перекрытия по Сёрсенсену-Дайсу.

Стохастическая сбалансированная по классам выборка

Второй проблемой сегментации малых целевых структур является крайний объёмный дисбаланс клас-

сов. Целевая область обычно занимает менее 1% воксельной сетки 256^3 . Оптимизация стандартных функций потерь на таких данных часто приводит к сходимости в вырожденный локальный минимум, при котором сеть однородно предсказывает класс «фон».

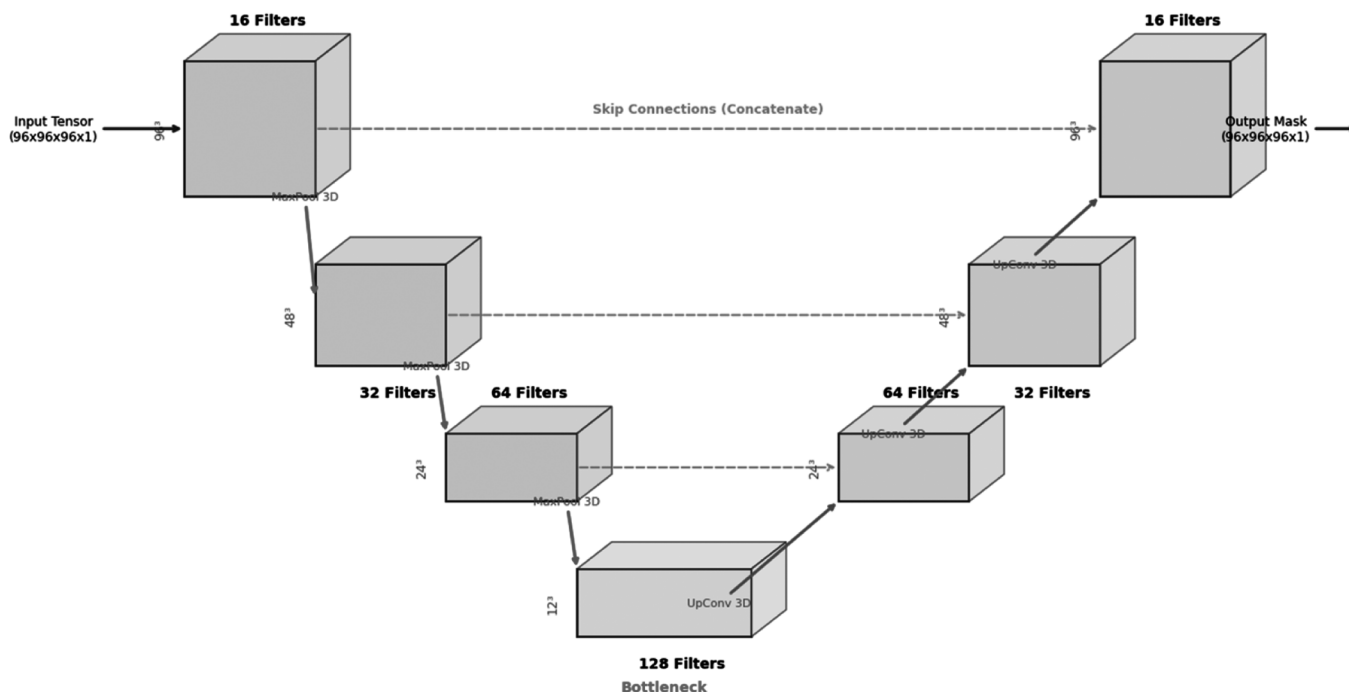
Для противодействия этому обучение выполняется с использованием стохастического алгоритма извлечения патчей (размер патча $P = 96^3$). Задаётся ограничение на вероятность выборки p :

$$P(\phi(P_k) = 1) \geq 0,6 \quad (3)$$

где $\phi(P_k)$ возвращает 1, если патч содержит целевые воксели переднего плана. Принудительное обеспечение того, чтобы 60 % всех генерируемых обучающих пакетов содержали полезный обучающий сигнал, математически предотвращает затухание градиентов. Также применяются интенсивные динамические трёхмерные пространственные аугментации (инверсии осей и ортогональные повороты) для искусственного расширения признакового пространства когорты из $N = 10$ субъектов.

Постановка эксперимента и протокол LOSO

Алгоритм проверялся на клиническом пилотном наборе из $N = 10$ высокоразрешающих T1-взвешенных MPT-сканов с использованием DLPFC в качестве сложной целевой структуры.



Note: Base filter count is restricted to 16, resulting in ~1.4M parameters to prevent overfitting on $N=10$.

Рис. 1. Схема предложенной облегчённой архитектуры трёхмерной U-Net. За счёт ограничения базового числа фильтров до 16 геометрическая прогрессия ($16 \rightarrow 32 \rightarrow 64 \rightarrow 128$) ограничивает общую параметрическую ёмкость примерно 1,4 млн параметров, выступая структурным регуляризатором против переобучения в режиме $N = 10$

Источник: разработано авторами

Учитывая ограниченный размер когорты, стандартное разбиение на обучающую и проверочную выборки внесло бы значительную дисперсию. Был реализован строгий протокол кросс-валидации с исключением одного субъекта (LOSO). Эксперимент состоял из $K = 10$ независимых фолдов. В каждом фолде один субъект полностью изолировался для инференса, а оставшиеся 9 использовались для обучения с нуля. Это гарантирует несмещённую оценку истинной обобщающей способности алгоритма на строго невиданной анатомии.

Количественные показатели

Основной метрикой точности семантической сегментации выступает коэффициент Сёренсена — Дайса (DSC).

Таблица 2.

Показатели сегментации (протокол LOSO)

Фолд	Размер обучающей выборки	Коэффициент Дайса (DSC)	Интерпретация
Случай 1	N=9	0.953	Высокая степень согласия
Случай 2	N=9	0.254	Плохое обобщение
Случай 3	N=9	0.969	Совпадение на экспертном уровне
Случай 4	N=9	0.448	Плохое обобщение
Случай 5	N=9	0.963	Высокая степень согласия
Случай 6	N=9	0.969	Совпадение на экспертном уровне
Случай 7	N=9	0.953	Высокая степень согласия
Случай 8	N=9	0.964	Высокая степень согласия
Случай 9	N=9	0.966	Высокая степень согласия
Случай 10	N=9	0.951	Высокая степень согласия
Среднее	—	0.838±0.24	—

Источник: разработано автором

Предложенный метод достиг среднего значения DSC $0,838 \pm 0,24$, что указывает на хорошее среднее качество, но значительный разброс между субъектами. Высокая точность сегментации ($DSC \geq 0,95$) была получена в 8 случаях из 10, демонстрируя, что модель способна выучивать точные анатомические представления в условиях крайнего дефицита данных. Для сложных кортикальных структур значения DSC, превышающие 0,90, обычно считаются соответствующими уровню согласия экспертов.

Однако два случая (Случай 2: 0,254, Случай 4: 0,448) показали серьёзное падение качества — это не случайно, а соответствует устойчивому режиму отказа. Визуаль-

ный анализ геометрии коры показывает, что у этих субъектов наблюдаются нетипичные паттерны кортикальной складчатости, недостаточно представленные в обучающем распределении ($N = 9$ на каждый фолд). Как следствие, выученное многообразие признаков не обладает достаточным покрытием редких анатомических вариаций, что служит эмпирическим свидетельством узкого места обобщения при малых данных, когда ограниченное разнообразие выборки сдерживает способность модели к экстраполяции на морфологии, выходящие за пределы обучающего распределения.

Динамика обучения и вычислительная эффективность

Несмотря на ограниченный размер набора данных, процесс обучения демонстрирует устойчивое оптимизационное поведение с плавной асимптотической сходимостью функции потерь на обучении.

Функция потерь на обучении (красный) демонстрирует плавную сходимость, тогда как метрика на проверке (синий) проявляет высокочастотную дисперсию, характерную для стохастического градиентного спуска при минимальном размере пакета ($B = 1$). Функция потерь на проверке демонстрирует характерную высокочастотную дисперсию, согласующуюся со стохастическим градиентным спуском в условиях экстремально малого размера пакета (размер пакета = 1). Хотя общий нисходящий тренд свидетельствует об эффективном обучении, сходимость функции потерь не повсеместно транслируется в обобщающую способность, что подтверждается вариативностью DSC на уровне отдельных фолдов.

В отличие от более тяжёлых трёхмерных архитектур сегментации, склонных к стремительному переобучению и нестабильной оптимизации в условиях малых данных, предложенная облегчённая 3D U-Net демонстрирует устойчивую сходимость по всем фолдам и достигает высокого качества в большинстве случаев. Не претендуя на абсолютную оптимальность, эти результаты указывают на то, что сокращение параметров действует как неявный структурный регуляризатор, повышая стабильность обучения и обеспечивая возможность научения в условиях острой нехватки данных, хотя и не разрешая полностью обобщение на редкие анатомические конфигурации.

В эксплуатационном отношении облегчённая сеть требует пикового объёма видеопамати всего лишь 4,3 Гб. С использованием алгоритма инференса на основе скользящего окна с динамическим дополнением полная реконструкция объёма достигалась за 36 секунд на пользовательском графическом процессоре NVIDIA Tesla T4, что делает данный конвейер высокопригодным для развёртывания на периферийных устройствах в режиме реального времени.

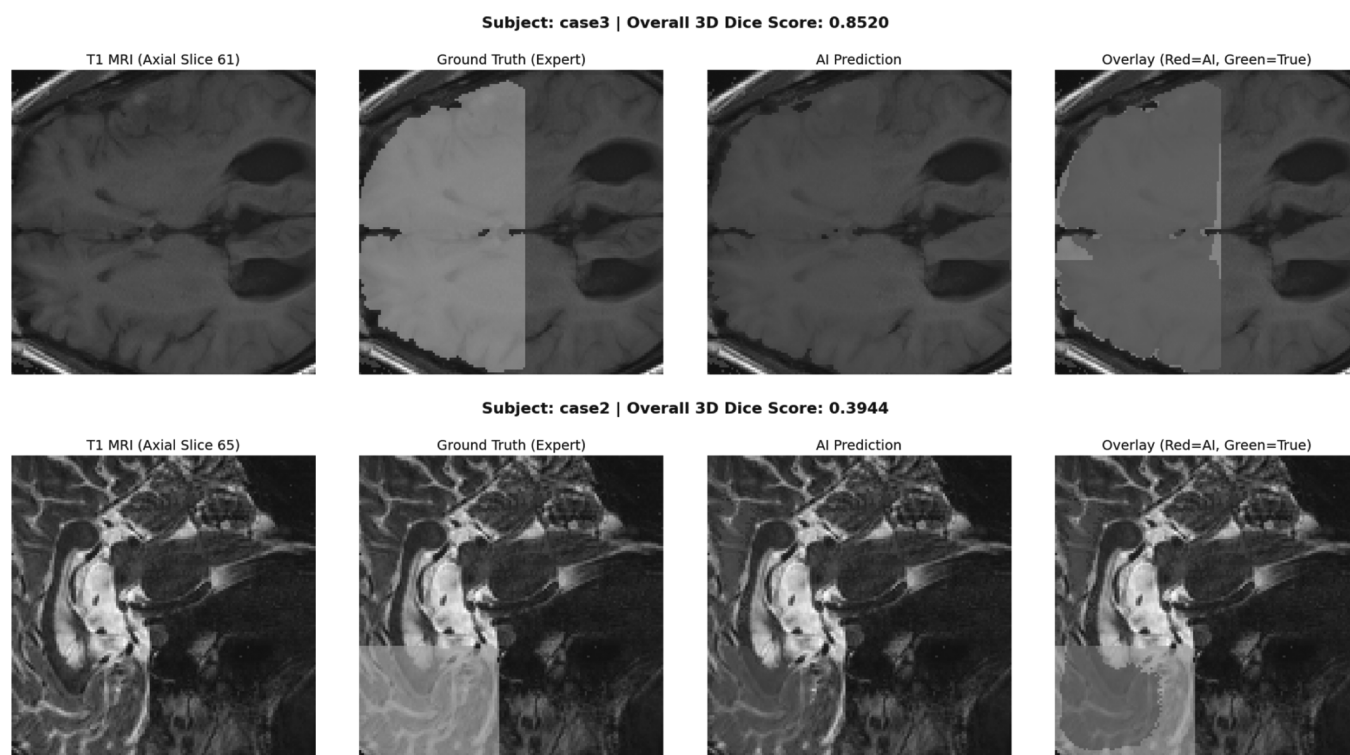


Рис. 3. Качественное сравнение точности сегментации. (Вверху) Успешная каноническая локализация (Случай 3, DSC 0,969), демонстрирующая точное следование границе серого и белого вещества. (Внизу) Основной режим отказа (Случай 2, DSC 0,254), где модель недоопределяет целевую область вследствие крайне нетипичной, прерывистой кортикальной складчатости, отсутствующей в обучающем многообразии

Источник: составлено авторами

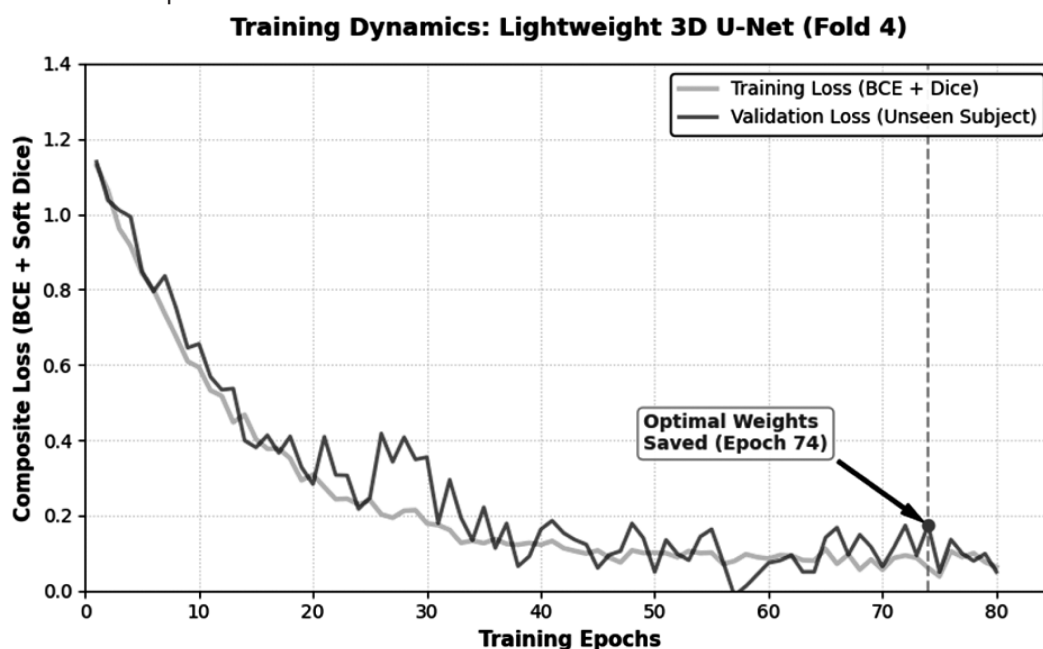


Рис. 4. Характерная динамика обучения для одного фолда LOSO

Источник: разработано авторами

В данной работе формализована и проверена эффективная по данным среда трёхмерной сегментации, адаптированная для режимов крайнего дефицита данных. Путём стратегического ограничения архитектурной ём-

кости трёхмерной U-Net до ~1,4 млн параметров и применения сбалансированного по классам стохастического алгоритма выборки мы математически ограничили ошибку обобщения. Протестированная посред-

ством строгого протокола LOSO на пилотной когорте из $N = 10$ субъектов, модель достигла пикового значения коэффициента Дайса 0,969, успешно извлекая сложные кортикальные топологии с нуля без использования массивных предварительно обученных наборов данных.

Данный подход предоставляет масштабируемую, вычислительно эффективную основу для развёртывания глубокого обучения в специализированных медицинских областях, где аннотированные данные изначально скудны.

ЛИТЕРАТУРА

1. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, «3D U-Net: learning dense volumetric segmentation from sparse annotation,» in Proc. MICCAI, Springer, 2016, pp. 424–432.
2. F. Milletari, N. Navab, and S. A. Ahmadi, «V-Net: Fully convolutional neural networks for volumetric medical image segmentation,» in Proc. 4th Int. Conf. 3D Vision (3DV), IEEE, 2016, pp. 565–571.
3. F. Isensee et al., «nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation,» Nature Methods, vol. 18, no. 2, pp. 203–211, 2021.
4. A. Hatamizadeh et al., «UNETR: Transformers for 3D medical image segmentation,» in IEEE/CVF Winter Conf. on Appl. of Computer Vision (WACV), 2022, pp. 574–584.
5. V.N. Vapnik, Statistical Learning Theory. New York: Wiley, 1998.

© Апан Кеннет Айнбуно (kenneth.89@mail.ru); Шичкина Юлия Александровна (strange.y@mail.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

СИСТЕМА АВТОМАТИЗИРОВАННОГО ВЗАИМОДЕЙСТВИЯ С ЗАКАЗЧИКОМ В УПРАВЛЕНИИ ПРОЕКТАМИ

SYSTEM OF AUTOMATED INTERACTION WITH CUSTOMER IN PROJECT MANAGEMENT

**N. Vasin
V. Shmakov**

Summary. The article presents an automated customer interaction system developed by the authors for the early stages of software engineering. The complex problem of the semantic gap between project participants under the conditions of limited computational resources of small development teams is considered. As a solution, the architecture of an intelligent conversational agent based on locally deployed Large Language Models (LLMs) is proposed. To ensure high performance of the system on affordable hardware, an advanced method of low-bit weight quantization (QTIP) is integrated, minimizing the distortion of the computational graph. The limitations of a fixed context window are overcome using the developed semantic memory extension module Smart Context based on the ChromaDB vector DBMS. A comparative analysis confirmed the high computational and functional efficiency of the proposed comprehensive solution.

Keywords: project management, requirements engineering, large language models, weight quantization, QTIP, Smart Context, semantic search, communication automation.

Васин Никита Андреевич

Санкт-Петербургский политехнический
университет Петра Великого
vasin.na@edu.spbstu.ru

Шмаков Владимир Эдуардович

Кандидат технических наук, доцент,
Санкт-Петербургский политехнический
университет Петра Великого
shmakov_ve@spbstu.ru

Аннотация. В статье представлена разработанная авторами система автоматизированного взаимодействия с заказчиком на ранних этапах проектирования программного обеспечения. Рассматривается комплексная проблема семантического разрыва между участниками проекта в условиях ограниченных вычислительных ресурсов малых команд разработчиков. В качестве решения предложена архитектура интеллектуального диалогового агента на базе локально развернутых больших языковых моделей (LLM). Для обеспечения высокой производительности системы на доступном оборудовании интегрирован передовой метод низкобитного квантования весов (QTIP), минимизирующий искажения вычислительного графа. Ограничения фиксированного контекстного окна преодолеваются с помощью разработанного модуля семантического расширения памяти Smart Context на базе векторной СУБД ChromaDB. Проведенный сравнительный анализ подтвердил высокую вычислительную и функциональную эффективность предложенного комплексного решения.

Ключевые слова: управление проектами, инженерия требований, большие языковые модели, квантование весов, QTIP, Smart Context, семантический поиск, автоматизация коммуникаций.

Введение и актуальность исследования

В современных условиях индустрии разработки программного обеспечения (ПО) критически важным фактором успешности реализации проектов выступает эффективность взаимодействия между заказчиком и исполнителем на ранних этапах жизненного цикла [1].

Для небольших ИТ-компаний, стартапов и самозанятых специалистов стадия сбора, анализа и формализации требований является наиболее уязвимой.

В условиях жестких ресурсных ограничений до 40–50 % рабочего времени проектных менеджеров и ведущих инженеров затрачивается на рутинные коммуникации, уточнение неструктурированных пожеланий и агрегацию разнородных исходных материалов.

В силу человеческого фактора это приводит к возникновению критических ошибок в техническом зада-

нии (ТЗ), исправлять которые на стадии кодирования и тестирования становится экспоненциально дороже.

Анализ отечественной и зарубежной практики ведения ИТ-проектов позволяет выделить ряд устойчивых системных проблем, возникающих на предпроектном этапе [1, 2]:

1. Проблема «семантического разрыва» — заказчики, как правило, не обладая глубокой технической экспертизой, формулируют свои пожелания на языке абстрактных бизнес-целей («хочу современный интернет-магазин») [1].

Разработчики же мыслят категориями конкретных функций, моделей данных и трудозатрат. Это несоответствие приводит к созданию неполных или технически неосуществимых спецификаций.

2. Хаотичность исходных артефактов — процесс передачи бренд-материалов, текстов, логотипов и прототипов не регламентирован.

Клиенты присылают файлы разного объема и форматов через разрозненные каналы связи, что требует от исполнителя колоссальных затрат на ручную систематизацию.

3. Эффект «расползания» границ проекта (scope creep) — в процессе первичных обсуждений требования постоянно меняются и дополняются без оперативного отражения их влияния на бюджет и сроки проекта, что порождает конфликты.
4. Субъективность первичных оценок — расчет стоимости и сроков выполнения работ на основе неструктурированных данных сильно зависит от интуиции менеджера, что несет в себе прямые финансовые риски для исполнителя.

Существующие программные решения (Jira, monday.com, Visure ALM) ориентированы на управление задачами внутри уже сформированной и обученной команды.

Они выполняют роль реактивных хранилищ информации, но не способны вести активный целенаправленный диалог с неподготовленным внешним пользователем (заказчиком), проводить автоматизированную валидацию требований на соответствие методологическим стандартам (SMART) и генерировать консолидированные проектные брифы.

Таким образом, возникает объективная научно-практическая потребность в разработке специализированного программного обеспечения промежуточного уровня (middleware), реализующего функции интеллектуального автоматизированного агента взаимодействия и управления проектами [3].

Методология и архитектурное проектирование системы

Система реализована как модульный серверный монолит, в котором доменные сервисы опосредуют взаимодействие пользовательского интерфейса с локальным вычислительным ядром LLM [4].

Основной способ сбора требований — диалог заказчика с ИИ-ассистентом в сочетании с загрузкой документов.

Поступающие высказывания естественного языка индексируются семантически и используются в RAG-пайплайне для структурирования проектного брифа, валидации и формирования уточняющих вопросов [3].

Ключевым функциональным блоком является модуль валидации требований. При обнаружении размытых,

двусмысленных или неколичественных формулировок, нарушающих критерии измеримости и определенности (SMART), система приостанавливает линейный опрос и генерирует таргетированный уточняющий вопрос [1].

Дополнительно система осуществляет классификацию и привязку сопутствующих проектных артефактов (схем, текстовых спецификаций, примеров интерфейсов).

Итоговым выходным документом стадии выявления требований является автоматически сформированный структурированный проектный бриф, включающий верифицированный перечень пользовательских историй (User Stories), матрицу трассируемости и предварительную оценку трудозатрат.

Высокие требования к конфиденциальности коммерческих данных клиентов исключают возможность трансляции проектных обсуждений во внешние облачные коммерческие API (OpenAI, Anthropic).

Это обуславливает необходимость развертывания LLM строго во внутреннем контуре организации (On-Premise) [5].

Для реализации серверной части авторами был выбран асинхронный веб-фреймворк FastAPI на языке Python 3.11+ [6].

В качестве основного хранилища структурированных данных, метаданных проектов и параметров авторизации используется реляционная СУБД PostgreSQL.

Для обеспечения долгосрочной ассоциативной памяти диалогового агента база данных расширена модулем pgvector, позволяющим оперировать векторными представлениями текстовых фрагментов.

Хранение неструктурированных бинарных объектов организовано в объектном S3-совместимом хранилище MinIO [7].

Решение проблемы ресурсных ограничений: метод квантования весов QTIP

Главным препятствием для локального внедрения современных высокоэффективных языковых моделей в малых командах является экстремально высокая потребность алгоритмов глубокого обучения в объеме видеопамати (VRAM) графических ускорителей.

Для преодоления данного барьера авторами был интегрирован передовой метод weight-only квантования весов нейронных сетей QTIP (Quantization with Trellises and Incoherence Processing), представленный на международной конференции NeurIPS 2024 [8].

Теоретический базис используемого метода QTIP (Quantization with Trellises and Incoherence Processing) направлен на радикальное снижение требований к видеопамяти при сохранении высокого качества работы языковой модели [8].

В рамках разрабатываемой платформы SRP данный метод применяется для оптимизации локального контура искусственного интеллекта (Ollama/FastAPI) и реализуется в два последовательных этапа:

Устранение «выбросов» данных (Incoherence Processing): На первом этапе веса большой языковой модели (LLM) проходят через специальное математическое преобразование (Адамара) [8].

Это позволяет «сгладить» аномальные всплески в структуре данных (outliers), которые обычно возникают при сильном сжатии моделей.

Благодаря этому модель, сжатая до 2–3 бит, не теряет логику рассуждений и не начинает «галлюцинировать» при анализе требований заказчика.

Параллельное решетчатое кодирование (Trellis Coded Quantization): На втором этапе веса эффективно упаковываются с помощью технологии решеток [8].

Метод QTIP оптимизирован под архитектуру CUDA, что позволяет осуществлять матрично-векторное умножение вида параллельно на графическом процессоре (GPU).

Настройка параметров сжатия (размерность плитки решетки через параметры `td_x`, `td_y` и калибровка таблицы `T_lut`) дополняется современным алгоритмом округления YAQA [8].

В отличие от базовых методов сжатия, YAQA минимизирует погрешности округления весов еще на этапе подготовки модели.

Семантическое расширение контекстной памяти посредством Smart Context

Второй фундаментальной проблемой применения локальных LLM в управлении проектами является жесткое ограничение длины контекстного окна.

Процесс детального сбора требований к крупной информационной системе может занимать недели и состоять из тысяч реплик, что неизбежно приводит к вытеснению ранних этапов беседы из оперативной памяти модели.

Для преодоления данного лимита авторами разработан и интегрирован программный модуль семан-

тического кэширования Smart Context, использующий в качестве технологической основы векторную СУБД ChromaDB [9].

Принцип работы модуля заключается в непрерывной фоновой векторизации и индексации поступающих сообщений [9].

С помощью специализированной модели эмбедингов (`omic-embed-text`) каждое сообщение трансформируется в плотный вектор фиксированной размерности и сохраняется в коллекции ChromaDB, привязанной к конкретному идентификатору проекта [9].

При отправке заказчиком новой реплики модуль Smart Context извлекает её вектор, выполняет косинусное сравнение близости со всей историей проекта и извлекает наиболее релевантные воспоминания.

Авторами были исследованы три стратегии инъекции извлеченного контекста в рабочий промпт модели:

1. Стратегия замещения (Replace oldest history): Из промпта удаляются наиболее старые в хронологическом порядке сообщения, а на их место бесшовно подставляются семантически близкие фрагменты из прошлых дней обсуждения [9].

Стратегия минимизирует риск переполнения контекста, однако может путать легковесные модели из-за нарушения естественного хода диалога.

2. Стратегия нижнего позиционирования (Add to Bottom): Локальная история сохраняется нетронутой, а извлеченные факты добавляются в самый конец системного промпта, будучи обернутыми в технический тег `[bracket header]` [9].

Это оказывает максимальное влияние на вектор генерации ответа ИИ, но повышает вероятность дублирования данных.

3. Стратегия настраиваемой глубины (Custom Depth): Позволяет осуществлять прецизионную вставку структурированных воспоминаний в промежуточные слои шаблона промпта через специализированный токен-заполнитель `{{comment}}` [9].

Экспериментальным путем для целей инженерии требований была выбрана комбинированная процентная стратегия (Use % Strategy), динамически вычисляющая долю замещаемого пространства контекста в зависимости от текущего наполнения буфера токенов.

Операционный контур управления проектом и модель данных

Функциональная архитектура разработанной платформы включает интегрированный контур операци-

онного управления, который обеспечивает сквозной жизненный цикл проекта — от первичного диалога до формирования пула рабочих задач.

Взаимодействие ключевых компонентов системы организовано на уровне трех основных сервисов:

Модуль ведения проектов и задач (Task Service): Обеспечивает стандартные процедуры создания, модификации и декомпозиции проектных задач.

В системе реализовано динамическое представление задач в формате интерактивной Канбан-доски с автоматической группировкой по рабочим спринтам, исполнителям и текущим статусам выполнения.

Сервис жизненного цикла и бизнес-логики (Workflow Service): Регламентирует правила валидации и переходов между состояниями задач и проекта в целом.

Компонент блокирует некорректные действия пользователей и обеспечивает жесткое методологическое ограничение: перевод ИТ-проекта на стадию непосредственной разработки технически невозможен до момента успешного прохождения автоматической проверки проектного брифа и фиксации базовых требований.

Асинхронная служба уведомлений (Notification Service): Осуществляет оперативное информирование участников команды через внутренние интерфейсы и SMTP-протоколы о критических событиях, включая распределение задач, приближение сроков сдачи (дедлайнов) и фиксацию изменений в спецификации требований.

Обеспечение сквозной трассируемости и целостности данных

Научно-практическая новизна архитектурного решения заключается в реализации механизма сквозной трассируемости «требование задача».

Каждая сформированная рабочая задача жестко ассоциирована со своим первоисточником — конкретным верифицированным блоком в схеме проектного брифа.

В случае внесения заказчиком или аналитиком корректировок в бриф, система автоматически идентифицирует все связанные задачи и помечает их флагом необходимости пересмотра (needs review), минимизируя риски потери информации при передаче требований в разработку.

Для предотвращения случайного затирания данных в процессе итерационного взаимодействия с ИИ внедрен программный механизм «запираания фактов» (AI Locks).

Подтвержденные или скорректированные аналитиком вручную разделы требований аппаратно блокируются от автоматической перезаписи алгоритмами RAG-генерации при последующих сессиях анализа документов.

Реляционная модель данных системы развернута в СУБД PostgreSQL и концептуально разделена на два изолированных уровня:

Коммуникационный контур: Агрегирует информацию о пользователях, проектах, сессиях чата, а также содержит коллекции векторов текстовых фрагментов и извлеченных фактов с прямыми ссылками на первоисточники в S3-хранилище.

Контур управления: Хранит сведения о задачах, спринтах, системных уведомлениях и истории переходов по бизнес-процессам.

Для оптимизации вычислительной нагрузки при построении Канбан-представлений на уровне базы данных применены составные b-tree индексы, снижающие время выборки информации при одновременной работе множества пользователей.

Логическая целостность данных поддерживается строгими ограничениями на уровне таблиц и каскадным удалением (ON DELETE CASCADE) связанных контекстных чанков и фактов при удалении родительского проекта.

Заключение

В ходе проведенного исследования сформировано комплексное научно-практическое решение, направленное на устранение коммуникационных потерь и автоматизацию процессов инженерии требований на ранних этапах жизненного цикла программного обеспечения.

Разработанная система, реализованная в виде модульного серверного монолита, обеспечивает эффективную интеграцию специализированных доменных сервисов (Chat, Brief, Embedding, AI) для организации контролируемого диалогового взаимодействия с заказчиком.

Использование локального контура больших языковых моделей на базе платформы Ollama в сочетании с разработанными механизмами семантического индексирования и RAG-пайплайнами позволило автоматизировать процессы структурирования проектных брифов, валидации поступающих формулировок по критериям SMART и генерации контекстных уточняющих вопросов [6].

Сравнительный анализ с существующими проектными трекерами (Jira , monday.com) показал, что создан-

ный программный прототип SRP успешно преодолевает их ключевые ограничения за счет перехода от реактивного хранения данных к проактивной модели ведения диалога [1].

Предложенный подход позволяет малым командам разработчиков и самозанятым специалистам ра-

дикально сократить временные затраты на рутинную координацию, снизить риски «расползания» границ проекта (scope creep) и обеспечить полную конфиденциальность обрабатываемых данных за счет изоляции ИИ-компонентов внутри периметра организации [3].

ЛИТЕРАТУРА

1. Wang Y., Luo L., Liu H. Bridging the semantic gap between customer needs and design specifications using user-generated content // IEEE Transactions on Engineering Management. — 2022. — Vol. 69, No. 4. — P. 1622–1634. — URL: <https://doi.org/10.1109/TEM.2020.3021698> (дата обращения: 20.05.2026).
2. Huang J. A soft prompts-based approach with pretrained language models for customer needs analysis. — Hong Kong University of Science and Technology. — 2024. — URL: <http://qlab.ieda.ust.hk/upload/paper170.pdf> (дата обращения: 15.05.2026).
3. Zhang Y., Li X., Wang J. User Intention Recognition and Requirement Elicitation Method for Conversational AI Services. — arXiv:2009.01509. — 2020. — URL: <https://arxiv.org/abs/2009.01509> (дата обращения: 18.05.2026).
4. Аникин Д.А. Анализ методов авторизации и аутентификации REST API // Информационные технологии и системы. — 2023. — С. 122–126.
5. Денисов Д.А. Искусственный интеллект как инструмент модерации контента // Современные информационные технологии. — 2024. — С. 19–22.
6. FastAPI Documentation [Электронный ресурс]. — Режим доступа: <https://fastapi.tiangolo.com/> (дата обращения: 22.01.2025).
7. Docker Documentation [Электронный ресурс]. — Режим доступа: <https://docs.docker.com/> (дата обращения: 24.01.2025).
8. Tseng A. QTIP: Quantization with Trellises and Incoherence Processing // NeurIPS 2024 Spotlight. Cornell-RelaxML Repository. — URL: <https://github.com/Cornell-RelaxML/qtip> (дата обращения: 31.05.2026).
9. Smart Context Module Documentation // Silly Tavern Docs. — URL: <https://docs.st.app/extensions/smart-context> (дата обращения: 31.05.2026).

© Васин Никита Андреевич (vasin.na@edu.spbstu.ru); Шмаков Владимир Эдуардович (shmakov_ve@spbstu.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

DOI 10.37882/2223-2966.2026.05-2.05

ОСОБЕННОСТИ ДООБУЧЕНИЯ ЭНКОДЕРНЫХ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ ДЛЯ РАБОТЫ С НОВОЙ ПРЕДМЕТНОЙ ОБЛАСТЬЮ ИСПОЛЬЗУЯ ФУНКЦИЮ ПОТЕРЬ TRIPLET LOSS

FEATURES OF FINE-TUNING ENCODER-BASED LARGE LANGUAGE MODELS FOR A NEW SUBJECT DOMAIN USING THE TRIPLET LOSS FUNCTION

**E. Vorobyov
N. Pikina
G. Rabotaev**

Summary. This article addresses the problem of domain adaptation of pre-trained language models to form a robust feature space for semantic search and relevant document retrieval (dense retrieval). The relevance of this study stems from the fact that the basic vector representations of universal models often fail to provide the required high quality of data extraction in narrow subject domains. The aim of this work is to develop and evaluate a fine-tuning model based on the Triplet Loss function, which adapts the geometry of frozen base embeddings using a small projection multilayer perceptron (MLP). In the course of the research, training sets were constructed as specialized triplets: «query (anchor) — relevant document (positive example) — non-relevant document (negative example)», and an algorithm for their effective contrastive learning was implemented. A comprehensive assessment of adaptation quality was conducted using standard information retrieval metrics (Recall, NDCG, MRR) along with a clear visualization of the feature space using t-SNE. The experimental results demonstrate that applying the Triplet Loss function significantly improves ranking quality and forms dense, well-separated semantic clusters. These findings confirm the high resource efficiency of the proposed method for embedding modification in the rapid construction of practical search systems.

Keywords: metric learning, Triplet Loss, embeddings, contrastive learning, domain adaptation, semantic search, feature space, dense retrieval, large language models, ranking.

Воробьев Ефим Алексеевич

Чебоксарский институт (филиал), Московский
политехнический университет, г. Чебоксары
efim-02@mail.ru

Пикина Наталия Евгеньевна

канд. пед. наук, доцент,
Чебоксарский институт (филиал), Московский
политехнический университет, г. Чебоксары
nataliapikina@mail.ru

Работаев Глеб Алексеевич

Чебоксарский институт (филиал), Московский
политехнический университет, г. Чебоксары
aylandbk@gmail.com

Аннотация. В статье рассматривается задача доменной адаптации предобученных языковых моделей для формирования устойчивого признакового пространства в задачах семантического поиска и отбора релевантных документов (dense retrieval). Актуальность исследования обусловлена тем, что базовые векторные представления универсальных моделей зачастую не обеспечивают требуемого высокого качества извлечения данных в узких предметных областях. Целью работы является разработка и оценка модели дообучения, основанной на применении функции потерь Triplet Loss, которая адаптирует геометрию замороженных базовых эмбеддингов с помощью небольшой проекционной полносвязной сети (MLP). В ходе исследования сформированы обучающие выборки в виде специализированных триплетов: «запрос (якорь) — релевантный документ (положительный пример) — нерелевантный документ (отрицательный пример)», и реализован алгоритм их эффективного контрастивного обучения. Проведена комплексная оценка качества адаптации с использованием стандартных метрик информационного поиска (Recall, NDCG, MRR) и наглядной визуализации пространства признаков методом t-SNE. Результаты проведенных экспериментов показывают, что применение функции Triplet Loss позволяет кардинально улучшить качество ранжирования и сформировать плотные, хорошо разделяемые семантические кластеры. Полученные результаты подтверждают высокую ресурсоэффективность предложенного метода модификации эмбеддингов для быстрого построения практических поисковых систем.

Ключевые слова: метрическое обучение, Triplet Loss, эмбеддинги, контрастивное обучение, доменная адаптация, семантический поиск, пространство признаков, dense retrieval, большие языковые модели, ранжирование.

Введение

В последние годы наблюдается стремительное развитие больших языковых моделей, которые находят широкое применение в задачах обработки

естественного языка. Особое место среди них занимают энкодерные модели, предназначенные для формирования векторных представлений текстовых данных. Такие модели позволяют эффективно решать задачи классификации, поиска и сравнения текстов за счёт использо-

вания эмбедингов, отражающих их семантическое содержание.

Несмотря на высокую универсальность предобученных языковых моделей, их применение в новой предметной области зачастую сопровождается снижением качества работы. Это обусловлено различиями в распределении данных, специфической терминологией и особенностями структуры текстов. В связи с этим возникает необходимость дообучения моделей с учётом характеристик целевой области, что позволяет адаптировать их к новым условиям использования.

Традиционные подходы к дообучению, основанные на решении задач классификации, не всегда обеспечивают формирование устойчивого признакового пространства, в котором корректно отражается семантическая близость объектов. В таких условиях перспективным является применение методов метрического обучения, направленных на оптимизацию расстояний между объектами в пространстве признаков.

Одним из наиболее эффективных инструментов метрического обучения является функция потерь Triplet Loss, позволяющая учитывать относительные различия между объектами. Использование триплетов вида «якорь — положительный — отрицательный» обеспечивает обучение модели таким образом, чтобы схожие тексты имели близкие векторные представления, а различающиеся — находились на значительном расстоянии друг от друга. Это особенно важно при адаптации моделей к новой предметной области, где требуется корректное отражение семантических различий.

Целью данной работы является исследование особенностей дообучения энкодерных больших языковых моделей для работы с новой предметной областью с использованием функции потерь Triplet Loss. Для достижения поставленной цели решаются задачи анализа подходов к адаптации языковых моделей, разработки метода формирования обучающих выборок, реализации алгоритма дообучения, а также оценки качества полученных эмбедингов.

Научная новизна работы заключается в применении подхода метрического обучения для адаптации энкодерных языковых моделей к новой предметной области с учётом особенностей данных. Практическая значимость определяется возможностью использования предложенного подхода в задачах анализа текстовой информации, требующих точного определения семантической близости объектов.

Материалы и методы

В рамках исследования решается задача адаптации энкодерной большой языковой модели к новой пред-

метной области с использованием методов метрического обучения. Основной целью является формирование такого признакового пространства, в котором семантически близкие объекты располагаются ближе друг к другу, а различающиеся — на большем расстоянии, что особенно важно при переносе модели на новые данные.

В качестве исходных данных используется набор текстовых объектов, относящихся к новой предметной области. Данные представлены в виде текстов или их фрагментов и сопровождаются разметкой, позволяющей определить семантическую близость между объектами. Перед использованием данные проходят этап предварительной обработки, включающий очистку, токенизацию и приведение к формату, совместимому с используемой языковой моделью.

В качестве основы используется предобученная энкодерная языковая модель, способная формировать векторные представления (эмбединги) входных текстов. Применение предобученной модели позволяет использовать уже сформированные знания о языке и сократить объём требуемых обучающих данных при адаптации к новой предметной области.

Дообучение модели осуществляется с учётом специфики новых данных, что позволяет повысить качество представлений и улучшить различимость объектов в рамках рассматриваемой задачи.

Обучение модели проводится с использованием триплетов — упорядоченных троек объектов вида «якорь (anchor), положительный пример (positive), отрицательный пример (negative)». Якорный и положительный объекты являются семантически близкими и относятся к одной категории, тогда как отрицательный объект отличается по смыслу и относится к другой категории.

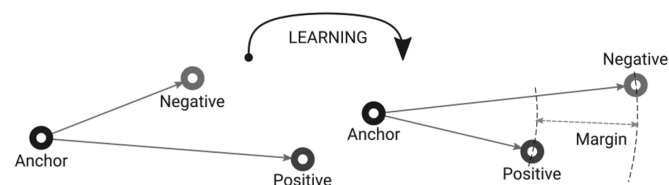


Рис. 1. Схема обучения модели с использованием триплетной функции потерь (Triplet Loss)

Формирование триплетов осуществляется на основе имеющейся разметки данных, что позволяет учитывать структуру предметной области и особенности распределения объектов.

Ключевым элементом дообучения является функция потерь Triplet Loss, позволяющая учитывать относительные расстояния между объектами в признаковом пространстве. Она обеспечивает обучение модели таким образом, чтобы расстояние между якорем и положи-

тельным примером было меньше, чем между якорем и отрицательным, с учётом заданного зазора (margin).

Функция потерь задаётся следующим образом:

$$L = \max(0, d(a, p) - d(a, n) + m)$$

где:

- a — якорный объект,
- p — положительный пример,
- n — отрицательный пример,
- $d(\cdot)$ — функция расстояния,
- m — параметр зазора.

Использование Triplet Loss позволяет эффективно адаптировать модель к новой предметной области за счёт явного учёта различий между объектами.

Дообучение модели осуществляется на сформированном наборе триплетов. На каждой итерации вычисляются эмбединги для элементов триплета, после чего рассчитывается значение функции потерь и выполняется обновление параметров модели с использованием градиентных методов оптимизации.

В отличие от классического fine-tuning, при котором оптимизируется функция классификации, использование Triplet Loss ориентировано на перестройку структуры признакового пространства, что делает данный подход более устойчивым при переносе модели на новые предметные области.

Оценка качества дообученной модели осуществляется на основе анализа расстояний между эмбедингами объектов. Рассматривается способность модели корректно сближать семантически близкие тексты и разделять различные. Для этого используются методы визуализации, а также метрики, основанные на расстояниях между объектами.

Литературный обзор

Современные исследования в области обработки естественного языка характеризуются активным использованием больших языковых моделей, в частности энкодерных архитектур, таких как BERT и его модификации. Данные модели позволяют формировать контекстно-зависимые представления текстов, что существенно повышает качество решения задач анализа текстовой информации. Отмечается, что использование трансформерных архитектур обеспечивает более точное отражение семантики текста по сравнению с традиционными методами [3, с. 31].

Несмотря на высокую эффективность предобученных моделей, их применение в новой предметной обла-

сти часто сопровождается снижением качества. Это связано с тем, что распределение данных в целевой области отличается от обучающего корпуса, что приводит к ухудшению обобщающей способности модели и требует её дополнительной адаптации.

Наиболее распространённым подходом является дообучение модели (fine-tuning), однако его использование не всегда обеспечивает формирование устойчивого признакового пространства. В этой связи особое внимание уделяется методам обучения представлений, в рамках которых *«представители одинаковых классов должны быть сгруппированы, а различных — разнесены»* [6, с. 10].

Данный принцип реализуется в методах метрического обучения, направленных на оптимизацию расстояний между объектами в пространстве признаков. Использование таких методов позволяет формировать более структурированное пространство представлений, что повышает качество задач сравнения и кластеризации данных [1, с. 48].

Одним из наиболее распространённых подходов метрического обучения является использование функции потерь Triplet Loss. В её основе лежит принцип, согласно которому *«якорный объект должен быть ближе к положительному примеру, чем к отрицательному»*. Это позволяет учитывать относительные различия между объектами и формировать более информативные эмбединги.

В то же время применение Triplet Loss связано с рядом ограничений. В частности, отмечается, что *«подход крайне чувствителен к выбору негативных примеров, что может приводить к нестабильности обучения»* [6, с. 11]. Это делает задачу формирования триплетов одним из ключевых факторов, влияющих на качество модели.

Дополнительные исследования показывают, что использование методов обучения представлений способствует улучшению структуры эмбедингового пространства и повышению качества разделения данных, особенно в условиях сложных и слаборазделимых классов.

Таким образом, анализ научной литературы показывает, что эффективная адаптация энкодерных языковых моделей к новой предметной области требует применения методов метрического обучения. Использование функции потерь Triplet Loss позволяет формировать устойчивое пространство признаков, в котором корректно отражается семантическая близость объектов, что делает данный подход перспективным для решения задач анализа текстовых данных.

Результаты

Экспериментальная часть исследования направлена на оценку эффективности применения контрастивного (метрического) обучения для доменной адаптации языковых моделей и формирования устойчивого признакового пространства.

Оценка проводилась комплексно: с одной стороны, замерялись количественные метрики качества на задаче поиска и отбора релевантных объектов (Dense Retrieval), с другой — проводился качественный визуальный анализ структуры сформированных эмбедингов.

В качестве метрик качества использовались стандартные показатели ранжирования: Recall@K (полнота — доля правильных ответов в топ-K выдачи), NDCG@K (нормализованная дисконтированная кумулятивная выгода, учитывающая позицию правильного ответа) и MRR (средний обратный ранг первого релевантного объекта).

На первом этапе было замерено базовое качество (baseline) исходной языковой модели без применения

доменной адаптации. Результаты поиска на «сырых» эмбедингах представлены в таблице 1.

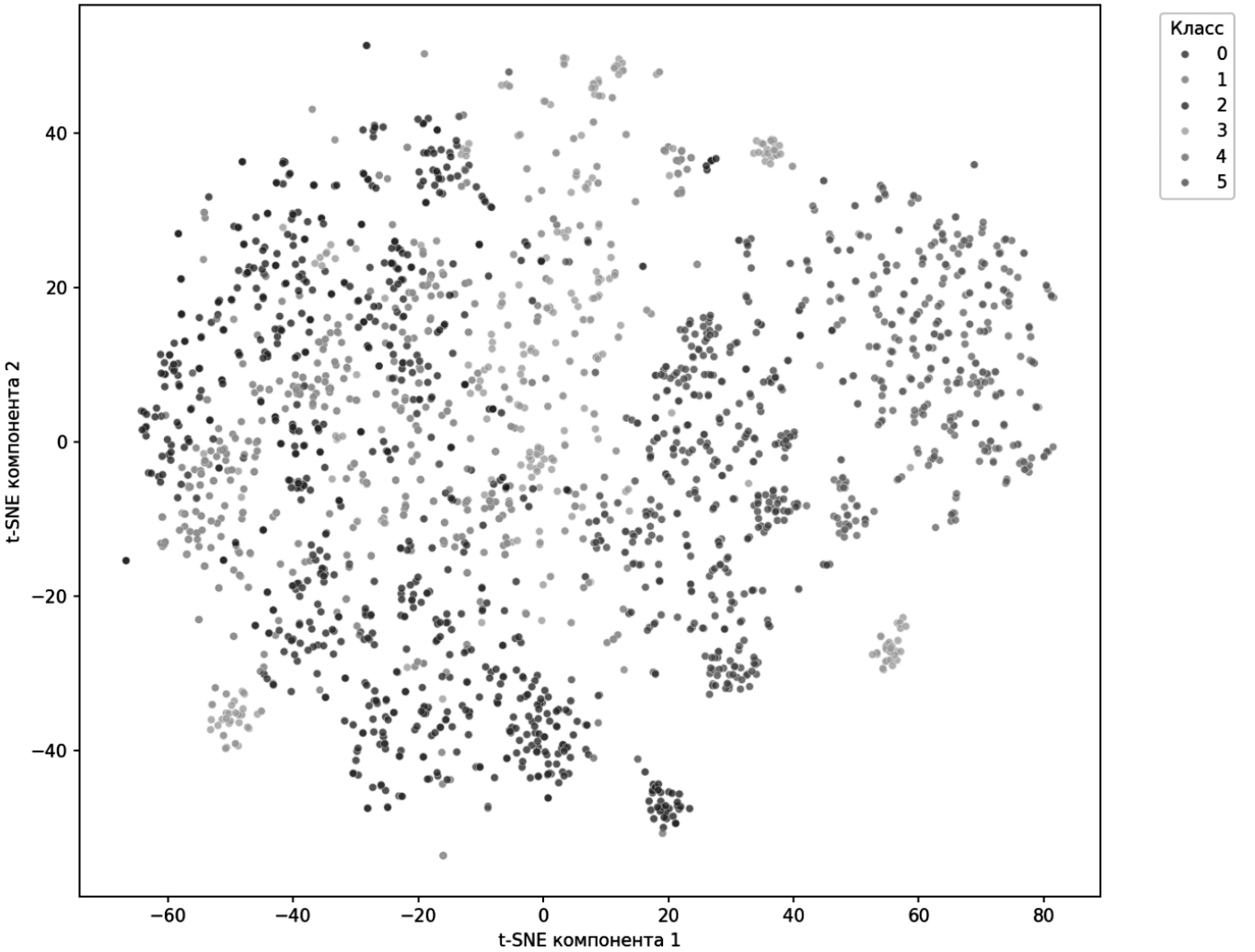
Таблица 1.

Базовые результаты ранжирования (до доменной адаптации)

Метрика	@10	@20	@50
Recall	0.0011	0.0014	0.0021
NDCG	0.0006	0.0007	0.0009
MRR	0.0005	0.0005	0.0005

Как видно из таблицы 1, базовая предобученная модель демонстрирует околонулевые метрики. Для понимания причин такого результата была выполнена визуализация исходного многомерного пространства признаков методом t-SNE (Рисунок 2).

На исходном распределении (Рисунок 2) наблюдается значительное зашумление и сильное взаимное перекрытие кластеров. Объекты из разных категорий слива-



ются воедино, что объясняет неспособность алгоритма находить релевантные соответствия: языковая модель «из коробки» не обладает разделяющей способностью для специфичных данных.

На следующем этапе была выполнена доменная адаптация. Поверх зафиксированных эмбеддингов обучалась полносвязная нейронная сеть (MLP) в контрастивном режиме, направленная на минимизацию расстояния между семантически связанными объектами (позитивными парами) и максимизацию расстояния до всех остальных объектов в обучающей выборке.

В процессе обучения наблюдалась устойчивая сходимость функции потерь. После завершения дообучения исходные данные были спроецированы в новое признаковое пространство, и процедура тестирования была повторена. Результаты представлены в таблице 2.

Сравнение таблиц 1 и 2 показывает радикальное улучшение качества работы модели. Метрика Recall@10 выросла на +0.6761, а Recall@50 достиг значения 0.8151

Таблица 2.

Результаты ранжирования после доменной адаптации (Contrastive Learning)

Метрика	@10	@20	@50
Recall	0.6773	0.7439	0.8151
NDCG	0.5140	0.5310	0.5452
MRR	0.4636	0.4682	0.4705

(в 81.5 % случаев релевантный объект теперь гарантированно находится в топ-50 выдачи). Значительный рост MRR (до 0.4636) свидетельствует о том, что искомый объект в среднем располагается на 2-й или 3-й позиции в итоговом списке кандидатов.

Природа такого резкого математического скачка наглядно демонстрируется на визуализации t-SNE, построенной для эмбеддингов после адаптации (Рисунок 3).

В отличие от исходного хаотичного состояния, геометрия нового признакового пространства кардинально

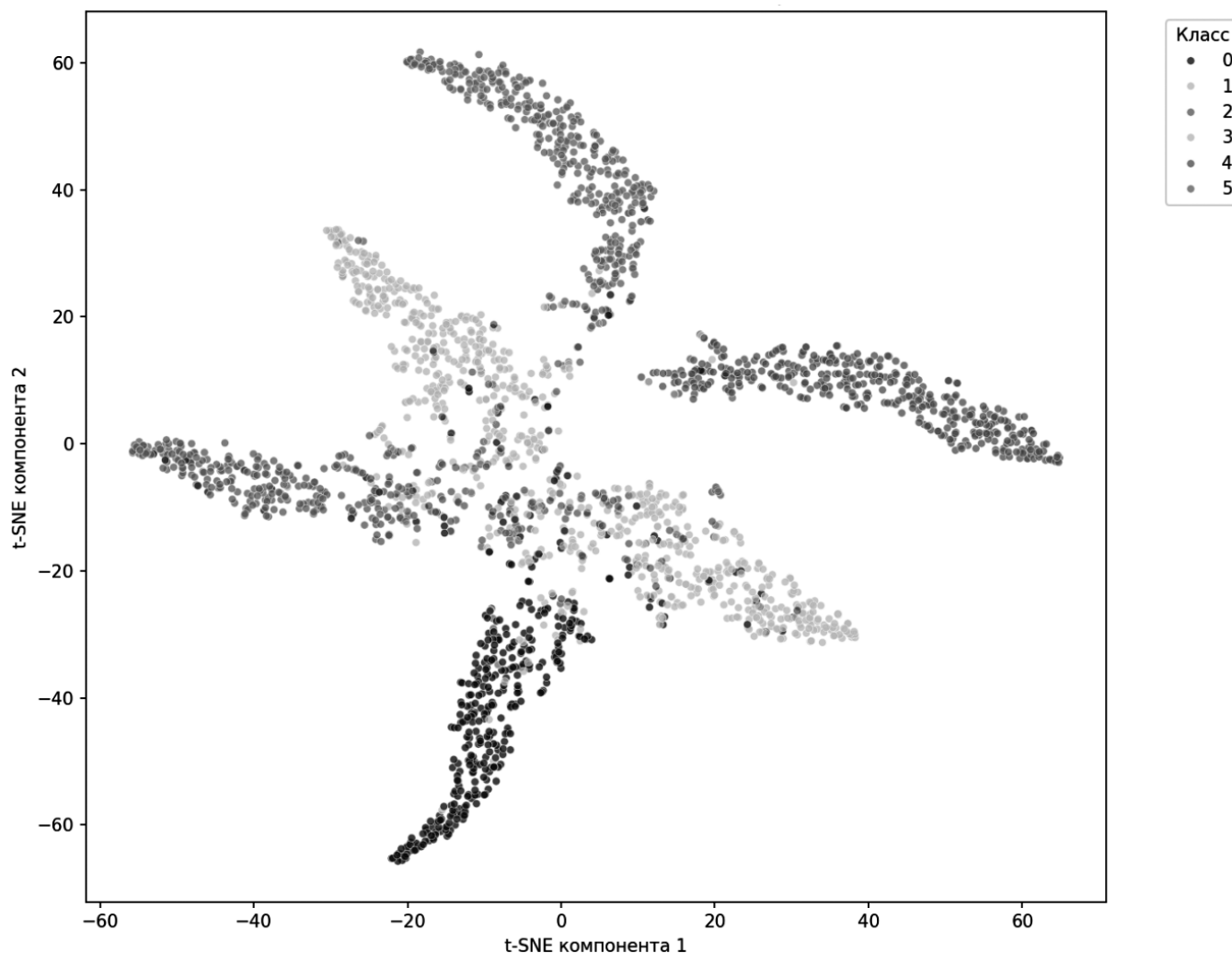


Рис. 3. Эмбеддинги после доменной адаптации

изменилась. Кластеры стали значительно более плотными, компактными и визуально изолированными друг от друга. Зоны перекрытия существенно уменьшились.

Таким образом, результаты эксперимента подтверждают, что применение методов метрического обучения позволяет не просто улучшить числовые метрики, но и глубоко модифицировать геометрию эмбедингов, формируя четко структурированное пространство с высокой степенью разделимости целевых классов.

Обсуждение

Полученные результаты ярко демонстрируют, что ключевым фактором успешной доменной адаптации является не столько исходная архитектура языковой модели, сколько выбранная парадигма её обучения. Колоссальный скачок метрик ранжирования (с околонулевых значений до $\text{Recall@50} > 0.81$) подтверждает, что базовая языковая модель обладала необходимым запасом знаний, однако извлечь их позволило только целенаправленное метрическое обучение.

Успех применения контрастивного подхода (основанного на принципах Triplet Loss и его современных модификациях) кроется в специфике процесса оптимизации. В отличие от традиционного обучения с учителем (например, через кросс-энтропию для задачи классификации), метрическое обучение оптимизирует относительные расстояния между объектами. Обучающий алгоритм не заставляет модель заучивать конкретные классы, а заставляет её выявлять скрытые семантические связи: штрафует за сближение разнородных текстов и поощряет сближение релевантных пар (запрос — документ). Именно такой процесс оптимизации приводит к перестройке признакового пространства, которую мы наблюдаем на t-SNE визуализациях.

Важным аспектом проведенного эксперимента является выбранная стратегия обучения. Адаптация проводилась путем заморозки основного «тела» энкодера (backbone) и обучения лишь небольшой проекционной полносвязной сети (MLP-головы). Этот подход к обучению имеет два фундаментальных преимущества: во-первых, он предотвращает эффект «катастрофического забывания» (catastrophic forgetting), сохраняя лингвистические знания базовой модели; во-вторых, он многократно снижает вычислительные затраты на расчет градиентов, делая процесс адаптации быстрым и ресурсоэффективным.

Тем не менее, процесс контрастивного дообучения имеет ряд существенных ограничений, которые необходимо учитывать на практике, например, зависимость от негативных примеров: Эффективность обучения кри-

тически зависит от того, с какими «негативами» сравнивается позитивная пара. Использование случайных негативов (random sampling) быстро перестает давать полезный градиент. Для поддержания стабильного процесса обучения требуется использование стратегий in-batch negatives (когда негативами выступают все остальные объекты в батче), что, в свою очередь, предъявляет высокие требования к объему видеопамати (требуется большой batch size). А также чувствительность к гиперпараметрам: Процесс метрического обучения крайне чувствителен к настройкам функции потерь (размер зазора/margin, температурный коэффициент). Некорректно подобранные параметры могут привести к коллапсу представлений (representation collapse), когда модель проецирует все объекты в одну точку пространства для минимизации ошибки.

В качестве направлений для дальнейших исследований целесообразно рассмотреть интеграцию в процесс обучения стратегий hard-negative mining (динамический подбор наиболее сложных для модели негативных примеров непосредственно в процессе тренировки). Также перспективным видится исследование процесса совместного обучения (joint training), при котором на поздних этапах адаптации происходит аккуратная разморозка верхних слоев базового энкодера, что может обеспечить еще более глубокое понимание доменной специфики.

Заключение

В ходе исследования была рассмотрена задача дообучения энкодерных больших языковых моделей для адаптации к новой предметной области с использованием функции потерь Triplet Loss. Показано, что данный подход позволяет формировать более качественные и семантически упорядоченные векторные представления текстовых данных, что особенно важно при работе с доменно-специфичной информацией. Применение Triplet Loss способствует сближению семантически близких объектов и увеличению расстояния между различающимися объектами в пространстве признаков, тем самым повышая разделимость классов и устойчивость получаемых эмбедингов. Полученные результаты подтверждают, что дообучение энкодерной модели с использованием Triplet Loss обеспечивает более эффективную адаптацию к новой предметной области и повышает ее пригодность для последующего применения в задачах семантического поиска, классификации и анализа текстов.

Таким образом, рассматриваемый подход может быть использован как перспективная основа для построения прикладных систем обработки специализированных текстовых данных.

ЛИТЕРАТУРА

1. Васильев А.В. Обзор методов метрического обучения с использованием глубоких нейронных сетей // Доклады Российской академии наук. Математика, информатика, процессы управления. — 2023. — Т. 514, № 2.
2. Кознов Д.В., и др. Calculating Similarity of Javadoc Comments // Programming and Computer Software. — 2024. — Т. 50, № 1. — С. 85–89.
3. Люткин Д.А., Поздняков Д.В., Соловьёв А.А. Transformer-Based Classification of User Queries for Medical Consultancy // Automation and Remote Control. — 2024. — Т. 85, № 3. — С. 297–308.
4. Новикова К.Н. Тенденции развития методов классификации объектов на основе модификаций контрастивного обучения // В сб.: Химия и инженерная экология — XXIII. — Казань, 2023. — С. 200–205.
5. Петрова Я.С. Анализ методов улучшения сходимости триплетной функции потерь в многометочной классификации // Моделирование, оптимизация и информационные технологии. — 2025. — Т. 13, № 4.
6. Пищев П.М. Обучение нейронных сетей на новых доменах и малых выборках // Информационные процессы. — 2024. — Т. 24, № 2. — С. 262–286.
7. Попов С.С., Лазарев М.С., Белавин В.А. Symbolic expression generation via variational auto-encoder // PeerJ Computer Science. — 2023.
8. Сурков А.В., и др. Application of large language models to solving differential equations // Smart Technologies, Systems and Applications. — Springer, 2024.
9. Юсупов В.С., Рахуба М.С., Фролов Е.В. Leveraging Geometric Insights in Hyperbolic Triplet Loss for Improved Recommendations // ACM Conference on Recommender Systems. — 2025.
10. Ефимов П.С., Бойцов Л.С., Арсланова Е.Р., Браславский П.И. The Impact of Cross-Lingual Adjustment of Contextual Word Representations on Zero-Shot Transfer // European Conference on Information Retrieval. — 2023.

© Воробьёв Ефим Алексеевич (efim-02@mail.ru); Пикина Наталия Евгеньевна (nataliapikina@mail.ru);

Работаев Глеб Алексеевич (aylandbk@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»

ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ БЕЗОПАСНОГО ИСПОЛЬЗОВАНИЯ ПЕРСОНАЛЬНЫХ ДАННЫХ В СИСТЕМАХ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: КОМПЛЕКСНЫЙ АНАЛИЗ ПРАВОВЫХ, ТЕХНИЧЕСКИХ И МЕТОДОЛОГИЧЕСКИХ АСПЕКТОВ

PROBLEMS AND PROSPECTS OF THE SAFE USE OF PERSONAL DATA IN ARTIFICIAL INTELLIGENCE SYSTEMS: A COMPREHENSIVE ANALYSIS OF LEGAL, TECHNICAL AND METHODOLOGICAL ASPECTS

S. Voronin

Summary. In the context of rapid digitalization and the widespread introduction of artificial intelligence (AI) technologies in critical areas of public life, the issue of personal data security has become of fundamental importance. This work is a comprehensive study of the privacy issues in machine learning (Machine Learning, ML), covering the legal conflicts arising from the application of Federal Law No. 152-FZ and international standards (GDPR, EU AI Act), as well as the technical vulnerabilities of modern deep learning architectures. The article provides a detailed taxonomy of attacks on AI models, including Model Inversion, Membership Inference, and gradient leakage in federated learning. Based on the analysis of security incidents in 2023–2025 and the results of benchmarking of Privacy-Enhancing Technologies (PETs), such as differential privacy, homomorphic encryption and secure multi-party computing (SMPC), scientifically based recommendations for building secure MLSecOps pipelines are formulated. The work aims to provide a comprehensive understanding of the current threat landscape and defense methods in the era of generative AI.

Keywords: artificial intelligence, machine learning, privacy, differential privacy, MLSecOps.

Воронин Сергей Анатольевич

Кандидат юридических наук, профессор, ЧУДПО
«Национальный научно-исследовательский институт»
info@voroninpravo.ru

Аннотация. В условиях стремительной цифровизации и повсеместного внедрения технологий искусственного интеллекта (ИИ) в критически важные сферы общественной жизни, вопросы обеспечения безопасности персональных данных приобретают фундаментальное значение. Настоящая работа представляет собой исчерпывающее исследование проблематики конфиденциальности в машинном обучении (Machine Learning, ML), охватывающее правовые коллизии, возникающие при применении Федерального закона № 152-ФЗ и международных стандартов (GDPR, EU AI Act), а также технические уязвимости современных архитектур глубокого обучения. В статье проводится детальная таксономия атак на модели ИИ, включая инверсию модели (Model Inversion), атаки на принадлежность (Membership Inference) и утечку градиентов в федеративном обучении. На основе анализа инцидентов безопасности 2023–2025 годов и результатов бенчмаркинга технологий сохранения приватности (Privacy-Enhancing Technologies, PETs), таких как дифференциальная приватность, гомоморфное шифрование и защищенные многосторонние вычисления (SMPC), формулируются научно обоснованные рекомендации по построению защищенных конвейеров MLSecOps. Работа призвана сформировать целостное представление о современном ландшафте угроз и методов защиты в эпоху генеративного ИИ.

Ключевые слова: искусственный интеллект, машинное обучение, конфиденциальность, дифференциальная приватность, MLSecOps.

Введение: Диалектика развития ИИ и права на неприкосновенность частной жизни

Современный этап научно-технического прогресса характеризуется беспрецедентным ростом возможностей систем искусственного интеллекта. Глубокие нейронные сети (Deep Neural Networks, DNN) и большие языковые модели (Large Language Models, LLM) демонстрируют способность решать задачи, ранее считавшиеся прерогативой человеческого интеллекта: от медицинской диагностики и управления автономным транспортом до генерации программного кода и креа-

тивного контента. Однако, фундаментом эффективности этих систем являются данные. Современные алгоритмы машинного обучения требуют для тренировки колоссальных объемов информации, значительная часть которой относится к категории персональных данных (ПДн) [1].

Возникает фундаментальное противоречие, которое можно охарактеризовать как конфликт между «информационным голодом» алгоритмов и правовыми принципами минимизации данных. С одной стороны, для повышения обобщающей способности (generalization)

и точности моделей необходимо использование максимально разнообразных и детальных датасетов. С другой стороны, законодательство о защите данных, как российское, так и международное, требует ограничения сбора информации конкретными, заранее определенными целями и удаления данных по достижении этих целей. Системы ИИ, способные запоминать (memorize) обучающие примеры и выявлять скрытые корреляции, создают новые векторы угроз, при которых традиционные методы обезличивания (деидентификации) становятся неэффективными [3].

Актуальность настоящего исследования обусловлена не только теоретической значимостью проблемы, но и резким ростом числа инцидентов, связанных с утечкой конфиденциальной информации через интерфейсы систем ИИ. Случаи непреднамеренного раскрытия коммерческой тайны и персональных данных сотрудников крупных корпораций через публичные сервисы генеративного ИИ в 2023–2024 годах продемонстрировали уязвимость существующего периметра безопасности. [5]. Более того, развитие методов атак на сами модели — таких как инверсия модели и атаки на принадлежность — переводит угрозу из плоскости инфраструктурной безопасности в плоскость математической безопасности алгоритмов.

Целью данной работы является систематизация накопленных знаний в области защиты данных в системах ИИ, выявление пробелов в текущем правовом регулировании и техническом инструментарии, а также разработка методологических подходов к обеспечению приватности на всех этапах жизненного цикла модели (ML Lifecycle).

Нормативно-правовое регулирование: Коллизии и гармонизация

Правовая среда, в которой функционируют системы ИИ, претерпевает значительные изменения. Традиционные нормы, разработанные для защиты статичных баз данных, адаптируются к динамичной природе алгоритмической обработки.

1. Российская правовая модель: Императивы ФЗ-152 и экспериментальные режимы

В Российской Федерации базовым нормативным актом является Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных». В контексте машинного обучения критическое значение приобретают несколько статей закона, применение которых к технологиям ИИ вызывает правовые дискуссии.

Во-первых, статья 10 устанавливает особый режим для обработки специальных категорий персональных

данных (сведения о расовой, национальной принадлежности, политических взглядах, состоянии здоровья, интимной жизни). Часть 4 данной статьи предписывает незамедлительное прекращение обработки таких данных, если устранены причины, вследствие которых она осуществлялась [8]. В парадигме классической информатизации это требование логично. Однако в контексте обучения нейронных сетей для медицинской диагностики возникает коллизия: модель может требовать периодического дообучения на исторических данных для поддержания актуальности (concept drift prevention). Удаление исходного датасета может сделать невозможным верификацию результатов работы модели и ее аудит, что противоречит принципам объяснимого ИИ (Explainable AI, XAI) [9].

Во-вторых, статья 16 (в новой редакции) закрепляет права субъектов при принятии решений на основании исключительно автоматизированной обработки. Данная норма коррелирует с международными стандартами, однако на практике сталкивается с проблемой «черного ящика» (black box problem). Если нейросеть принимает решение об отказе в кредите, оператор обязан разъяснить порядок принятия решения. Техническая сложность интерпретации весов глубоких сетей делает выполнение этого требования нетривиальной задачей [10].

Для преодоления барьеров, сдерживающих внедрение инноваций, государство использует механизм экспериментальных правовых режимов (ЭПР). В соответствии с Постановлением Правительства РФ от 18.07.2023 № 1164, установлен ЭПР в сфере цифровых инноваций по направлению медицинской деятельности. Данный режим, действующий в течение трех лет, позволяет участникам эксперимента (разработчикам медицинских ИИ-сервисов) обрабатывать сведения о состоянии здоровья и диагнозах граждан с применением технологий больших данных и ИИ в упрощенном порядке, при условии обеспечения надежного обезличивания [11].

Аналогичные инициативы реализуются в транспортной сфере. Постановление Правительства об установлении ЭПР для эксплуатации высокоавтоматизированных транспортных средств позволяет тестирование беспилотных такси в Москве (район Ясенево), Иннополисе и на федеральной территории «Сириус». Видеопотоки с камер беспилотников неизбежно содержат биометрические данные пешеходов и водителей других транспортных средств. ЭПР легализует эту обработку, устанавливая при этом жесткие требования к страхованию ответственности и кибербезопасности [12]. Такой подход позволяет нарабатывать правоприменительную практику в контролируемых условиях («песочницах») до внесения изменений в федеральное законодательство.

2. Европейский опыт: От GDPR к AI Act

Европейский союз демонстрирует наиболее проработанный подход к регулированию цифровой сферы, который оказывает влияние на глобальные стандарты (эффект Брюсселя).

GDPR (General Data Protection Regulation):

Регламент GDPR фокусируется на защите прав субъектов данных. Ключевые принципы — ограничение целью (purpose limitation) и минимизация данных (data minimization) — вступают в конфликт с практикой Big Data. Статья 22 GDPR предоставляет субъекту право не подвергаться решению, основанному исключительно на автоматизированной обработке, если оно порождает правовые последствия. Это создает существенные ограничения для полностью автономных систем скоринга или найма персонала [10].

EU AI Act (Акт об искусственном интеллекте):

Принятый AI Act представляет собой первый в мире комплексный закон, регулирующий непосредственно технологии ИИ. В отличие от GDPR, он использует риск-ориентированный подход, классифицируя системы ИИ по уровню потенциального вреда:

- 1. Неприемлемый риск (Unacceptable risk): системы, использующие сублиминальные техники манипуляции сознанием, эксплуатирующие уязвимости групп населения или реализующие социальный рейтинг (social scoring). Такие системы полностью запрещены.
- 2. Высокий риск (High risk): системы, применяемые в критической инфраструктуре, образовании, правоохранительной деятельности, управлении миграцией. Для них установлены строгие требования: высокое качество данных (data governance), ведение технической документации, прозрачность для пользователей, человеческий надзор (human oversight) и высокий уровень кибербезопасности [13].
- 3. Ограниченный риск (Limited risk): системы, взаимодействующие с людьми (чат-боты), системы распознавания эмоций. Основное требование — прозрачность (пользователь должен знать, что общается с машиной).
- 4. Минимальный риск (Minimal risk): большинство систем ИИ (спам-фильтры, видеоигры), для которых не вводятся дополнительных ограничений [15].

Гармонизация и Digital Omnibus 2025:

Взаимодействие GDPR и AI Act создает сложный ландшафт требований (compliance landscape). Напри-

мер, выполнение требования GDPR об удалении данных («право на забвение») может потребовать применения сложных технологий машинного разучивания (Machine Unlearning), чтобы удалить влияние данных из уже обученной модели. В 2025 году Еврокомиссия предложила пакет инициатив «Digital Omnibus», направленный на упрощение правил. Одной из ключевых новелл является предложение уточнить понятие «легитимного интереса» для разработчиков ИИ, что может упростить использование персональных данных для тренировки моделей при соблюдении определенных гарантий. Однако правозащитники опасаются, что это может привести к ослаблению защиты приватности [16].

Таблица 1.
Сравнительный анализ регуляторных подходов (РФ vs ЕС)

Критерий сравнения	ФЗ-152 «О персональных данных» (РФ)	GDPR / EU AI Act (ЕС)
Объект регулирования	Персональные данные как таковые, процесс их обработки оператором.	Данные (GDPR) и технологии ИИ как продукты (AI Act).
Подход к рискам	Формальный (по категориям данных и типам угроз в модели угроз ФСТЭК).	Риск-ориентированный (градация систем ИИ по степени опасности для прав человека).
Автоматизированные решения	Право на обжалование, обязанность разъяснения порядка принятия решения (ст. 16).	Право на человеческое вмешательство, запрет определенных видов решений, требования объяснимости.
Инновационные механизмы	Экспериментальные правовые режимы (ЭПР) с территориальным и отраслевым ограничением.	Регуляторные песочницы (Regulatory Sandboxes), предусмотренные AI Act для всего рынка ЕС.
Оценка воздействия	Требуется в рамках определения уровня защищенности ИСПДн.	Обязательная оценка воздействия на защиту данных (DPIA) для высокорисковых технологий.
Штрафные санкции	Фиксированные штрафы (КоАП РФ),кратно увеличивающиеся при повторных нарушениях.	Оборотные штрафы (до 4% глобального оборота по GDPR, до 7 % по AI Act).

Таксономия угроз приватности в машинном обучении

Понимание механизмов атак на системы ИИ является необходимым условием для построения эффективной защиты. Угрозы приватности в ML отличаются от классических киберугроз тем, что они эксплуатируют не программные ошибки (баги), а фундаментальные статистические свойства алгоритмов обучения.

1. Атаки инверсии модели (Model Inversion Attacks)

Атаки инверсии модели (MIA) направлены на реконструкцию входных данных или их чувствительных признаков на основе доступа к выходам модели. В сценарии «черного ящика» (Black-Box), где злоумышленник имеет доступ только к API прогнозирования (например, MLaaS), атака строится на итеративном подборе входа.

Математическая формулировка:

Злоумышленник стремится найти входной вектор x^* , который максимизирует апостериорную вероятность целевого класса c , заданную моделью M :

$$x^* = \underset{x}{\operatorname{argmax}} P(M(x) = c)$$

Процесс часто реализуется через градиентный спуск в пространстве входов. Начиная со случайного шума или усредненного изображения, злоумышленник изменяет пиксели в направлении градиента целевого класса.

Примеры и последствия:

Исследования показывают, что такие атаки особенно эффективны против систем распознавания лиц. Fredrikson et al. продемонстрировали возможность восстановления узнаваемого изображения лица человека по модели, обученной предсказывать имена. В 2024–2025 годах появились более совершенные методы, использующие генеративно-состязательные сети (GAN) в качестве априорного знания (priors) для генерации более реалистичных реконструкций, способных обходить простые фильтры. [18] В медицине это грозит восстановлением генотипа пациента по результатам фармакогенетической модели [20].

2. Атаки на принадлежность к обучающей выборке (Membership Inference Attacks)

Атака на принадлежность (Membership Inference Attack, MI) ставит своей целью определить, использовались ли данные конкретного индивида при обучении целевой модели. Успешная атака нарушает конфиденциальность, так как сам факт участия в выборке может быть чувствительным (например, база пациентов наркологического диспансера).

Механизм реализации:

Фундаментальная причина уязвимости к MI — переобучение (overfitting). Модель, как правило, «увереннее» (выдает более высокие вероятности для правильного класса, меньшую энтропию предсказания) на данных, которые она видела в процессе обучения, чем на новых тестовых данных.

Злоумышленник обучает набор «теневых моделей» (shadow models) M_{shadow} на данных, распределенных аналогично целевым. Затем он обучает бинарный классификатор (attack model) на выходах теневых моделей, который учится различать паттерны ответов для членов (IN) и не-членов (OUT) обучающей выборки [21].

Эволюция атак (2024-2025):

Новейшие исследования демонстрируют методы атак, не требующие обучения теневых моделей (LiRA — Likelihood Ratio Attack), которые сравнивают вероятность записи на целевой модели с вероятностью на ряде референсных моделей. Также появились атаки Range Membership Inference (RaMIA), определяющие принадлежность данных к определенному интервалу значений, что расширяет спектр угроз для регрессионных моделей [23].

3. Утечка данных в распределенных системах (Gradient Leakage)

Федеративное обучение (Federated Learning, FL) позиционируется как приватный метод обучения, так как данные не покидают устройства пользователя. Однако обмен градиентами (обновлениями весов) несет в себе риски.

Исследования 2025 года показывают, что хотя простые методы защиты (клиппинг градиентов) могут затруднить атаку, использование вспомогательной информации (Batch Normalization statistics) позволяет восстанавливать изображения с высоким разрешением даже при наличии защиты [25].

4. Извлечение данных из генеративных моделей (LLM и Diffusion Models)

Большие языковые модели (LLM), такие как GPT-4, Llama 3, и диффузионные модели (Stable Diffusion), подвержены риску запоминания обучающих данных (verbatim memorization).

Атаки на извлечение (Training Data Extraction):

Исследователи Carlini и др. разработали методы, позволяющие извлекать из моделей конкретные фрагменты текста: PII, ключи API, фрагменты кода. Атака Neural Phishing (2024) демонстрирует новый вектор: злоумышленник может внедрить в данные пре-тренинга (pre-training) специальные «якоря» (poisoned samples), которые заставляют модель запомнить секретную информацию (например, данные кредитных карт), поданную ей позже на этапе дообучения (fine-tuning). Эффективность извлечения таких секретов может достигать 50 %, что ставит под угрозу безопасность дообучения моделей на корпоративных данных [4].

Эмпирический анализ инцидентов безопасности (2023–2025)

Анализ реальных инцидентов позволяет перейти от теоретических моделей угроз к практической оценке рисков для бизнеса и государства.

1. Кейс Samsung: Утечка через MLaaS (2023)

В 2023 году инженеры полупроводникового подразделения Samsung использовали чат-бот ChatGPT для оптимизации исходного кода и конвертации заметок совещаний в презентации. В результате конфиденциальный код (связанный с измерением выхода годных чипов) и стенограммы стратегических совещаний были переданы на серверы OpenAI. Поскольку политика использования сервиса по умолчанию предполагала использование пользовательских данных для дообучения моделей, эта информация стала частью обучающего корпуса. Теоретически, другой пользователь мог бы извлечь эти фрагменты через специально сконструированные запросы (prompt engineering).

Последствия: Samsung ввела полный запрет на использование генеративного ИИ на корпоративных устройствах и инициировала разработку собственной внутренней модели. Инцидент подчеркнул риск использования публичных облачных сервисов ИИ для обработки чувствительных корпоративных данных (Shadow AI) [5].

2. Атака Midnight Blizzard на Microsoft (2024)

Группировка Midnight Blizzard (ассоциируемая с государственными структурами) скомпрометировала корпоративную почту Microsoft. Хотя атака началась с password spraying на тестовый аккаунт, злоумышленники целенаправленно искали информацию о работе систем ИИ, переписку инженеров по безопасности и данные о механизмах защиты моделей. Это свидетельствует о том, что интеллектуальная собственность в сфере ИИ и данные о его архитектуре становятся приоритетными целями для промышленного и государственного шпионажа. Кроме того, исследователи Wiz обнаружили утечку 38 ТБ данных Microsoft AI через GitHub-репозиторий из-за неправильно настроенного SAS-токена, что демонстрирует, как ошибки конфигурации (misconfiguration) остаются доминирующим вектором утечек даже в высокотехнологичных компаниях [31].

3. Уязвимости Google Gemini (2024)

В 2024 году были зафиксированы случаи, когда модель Google Gemini выдавала пользователям ответы, содержащие косвенные признаки обучающих данных, или демонстрировала неадекватное поведение при определенных промптах, что вызывало репутационные риски.

Также исследователи безопасности выявили возможности для косвенных инъекций промптов (Indirect Prompt Injection), когда модель, обрабатывая веб-страницу или документ, выполняла скрытые инструкции злоумышленника, что могло привести к эксфильтрации данных пользователя [34].

Технологии сохранения приватности (PETs): Бенчмаркинг и эффективность

Для минимизации описанных рисков применяется спектр технологий Privacy-Enhancing Technologies (PETs). Выбор конкретной технологии всегда представляет собой компромисс между приватностью, точностью (utility) и производительностью (performance).

1. Дифференциальная приватность (Differential Privacy, DP)

Дифференциальная приватность обеспечивает математическую гарантию того, что вывод алгоритма не изменится существенно при добавлении или удалении одной записи из обучающей выборки.

Алгоритм DP-SGD (Differentially Private Stochastic Gradient Descent):

Инструменты и бенчмарки (2024-2025):

Основные библиотеки: Opacus (для PyTorch) и TensorFlow Privacy.

Сравнительные тесты показывают, что Opacus обеспечивает более эффективное использование памяти благодаря векторизованному вычислению градиентов по образцам (per-sample gradients), однако обучение с DP все равно замедляет процесс в 2–5 раз по сравнению с обычным обучением.

Trade-off: для достижения высокой точности (например, >90 % на CIFAR-10) при строгой приватности требуются огромные вычислительные ресурсы или использование предобученных на публичных данных моделей с последующим DP-дообучением (DP-Fine-Tuning). Методы LoRA (Low-Rank Adaptation) в сочетании с DP показывают лучшие результаты для LLM, позволяя сохранить качество генерации [37].

2. Гомоморфное шифрование (Homomorphic Encryption, HE)

HE позволяет выполнять операции сложения и умножения над зашифрованными данными. Это «Святой Грааль» для аутсорсинга вычислений.

Состояние технологии:

Схемы CKKS (Cheon-Kim-Kim-Song) поддерживают приближенные вычисления над вещественными числами, что идеально подходит для нейросетей. Однако полносвязные операции (умножение матриц) в HE крайне ресурсоемки.

Бенчмарки: инференс простой сверточной сети (CNN) на зашифрованном изображении (MNIST) может занимать секунды на CPU. Использование аппаратных ускорителей (FPGA, GPU) в 2025 году позволило снизить задержку до субсекундного уровня для простых моделей, но для глубоких сетей (ResNet-50) задержки все еще измеряются десятками секунд или минутами, что ограничивает применение HE сценариями, не требующими реального времени (например, медицинский анализ) [40].

3. Защищенные многосторонние вычисления (SMPC)

SMPC позволяет участникам вычислить функцию, не раскрывая свои входы друг другу. Данные разбиваются на «секретные доли» (secret shares).

Фреймворки:

CrypTen (Meta) и PySyft (OpenMined) предоставляют абстракции тензоров, работающих поверх протоколов SMPC.

Сравнение с HE: SMPC вычислительно намного легче, чем не (на порядки), но требует интенсивного сетевого обмена данными между участниками (communication heavy). В тестах на классификацию изображений (ResNet-18) CrypTen демонстрирует время инференса около 2.5 секунд, что приемлемо для многих задач [44].

4. Синтетические данные (Synthetic Data)

Генерация синтетических данных с использованием генеративных моделей, обученных с DP, позволяет создавать безопасные двойники (digital twins) чувствительных датасетов.

Оценка приватности:

Критически важно не просто генерировать данные, но и проверять их на переобучение. Метрики «Distance to Closest Record» (DCR) и проведение имитационных атак на принадлежность (simulated MIA) являются стандартом качества синтетических данных. Исследования показывают, что использование DP-GAN или DP-Diffusion моделей позволяет снизить риск успешной атаки MIA до уровня случайного угадывания, сохраняя при этом статистическую полезность данных для обучения классификаторов [48].

Методологический фреймворк MLSecOps

Обеспечение безопасности ИИ требует интеграции защитных мер в процессы разработки и эксплуатации (DevOps). Парадигма MLSecOps объединяет специалистов по Data Science, безопасности и операциям.

1. Жизненный цикл безопасной разработки (Secure AI SDLC)

1. Проектирование (Design): Определение требований приватности. Выбор архитектуры (централизованная vs федеративная). Моделирование угроз (Threat Modeling) с учетом специфики ML (атаки на данные, атаки на модель).
2. Сбор данных (Data Ingestion): санитизация данных, удаление PII или их токенизация. Использование методов Local Differential Privacy (LDP) на этапе сбора. Data Versioning (DVC) для отслеживания происхождения данных (provenance) [52].
3. Разработка и обучение (Development & Training): обучение в изолированных средах («песочницах»). Использование проверенных библиотек (защита от Supply Chain attacks). Применение DP-SGD для защиты весов модели.
4. Тестирование (Testing): проведение Adversarial Testing (проверка устойчивости к состязательным примерам). Аудит на предмет предвзятости (Fairness) и утечек данных (Privacy Auditing с помощью инструментов типа TensorFlow Privacy).
5. Развертывание (Deployment): защита API (Rate limiting, аутентификация). Мониторинг входов на предмет аномалий и атак инверсии. Использование TEE (Trusted Execution Environments, например, Intel SGX) для защиты модели в облаке.
6. Мониторинг и реагирование: отслеживание дрейфа данных (Data Drift) и попыток атак. План реагирования на инциденты, включающий процедуры отзыва модели и машинного разучивания (Machine Unlearning) скомпрометированных данных [54].

2. Стандартизация и комплаенс

Для построения процессов MLSecOps необходимо опираться на международные стандарты:

- NIST AI RMF (Risk Management Framework): предлагает структуру Govern-Map-Measure-Manage для управления рисками. Профиль NIST SP 800-218A детализирует практики безопасной разработки для генеративного ИИ [56].
- ISO/IEC 42001:2023: стандарт системы менеджмента ИИ, обеспечивающий основу для сертификации. Он требует документирования всех решений, касающихся данных и алгоритмов, что критически важно для соблюдения принципа подотчетности (Accountability) в GDPR и AI Act [59].

Таблица 2.

Детальное сравнение технологий сохранения приватности (PETs) по ключевым метрикам

Технология	Основной принцип	Гарантии безопасности	Влияние на производительность	Влияние на точность (Utility)	Типовые сценарии использования	Зрелость технологии (TRL)
Дифференциальная приватность (DP)	Добавление калиброванного шума к данным или градиентам.	Теоретико-информационная (не зависит от вычислительной мощности атакующего).	Замедление обучения (2х-5х), увеличение потребления памяти.	Снижение точности, зависимое от бюджета приватности. Требуется тонкой настройки.	Сбор телеметрии (Apple, Google), обучение моделей на медданных, публикация статистики.	Высокая (промышленное внедрение).
Гомоморфное шифрование (HE)	Вычисления над шифротекстом без расшифровки.	Криптографическая (стойкость зависит от длины ключа).	Экстремально высокое (замедление в 103–104 раз без ускорителей).	Нет потери точности (для точных схем) или незначительная (для приближенных, как CKKS).	Inference-as-a-Service для высокочувствительных данных (геномика, финансы), PSI.	Средняя (активные пилоты, требует железа).
SMPC (Secure Multi-Party Computation)	Разделение секрета (Secret Sharing) между участниками.	Криптографическая (при условии честности большинства или всех участников).	Высокая нагрузка на сеть (Communication bottleneck). Вычисления быстрее HE.	Нет потери точности.	Совместное обучение моделей конкурентами (банки, страховые), агрегация данных.	Средняя/Высокая (используется в финтехе).
Федеративное обучение (FL)	Локальное обучение, агрегация обновлений.	Приватность данных (данные не покидают устройство). Уязвимо к атакам на градиенты без доп. защиты.	Нагрузка на клиентские устройства, сложная оркестрация, сетевые задержки.	Возможно снижение точности из-за non-IID данных (разнородность распределения).	Мобильные устройства (T9, предиктивный ввод), IoT, медицинские консорциумы.	Высокая.
Синтетические данные	Генерация искусственных выборок.	Зависит от качества генератора и проверки на переобучение (эмпирическая).	Нет влияния на инференс. Затраты на обучение генератора.	Зависит от fidelity (точности воспроизведения распределения). Риск потери сложных связей.	Тестирование ПО, обучение моделей при дефиците данных, Data Sharing.	Высокая.

- Guidelines for Secure AI System Development: совместное руководство агентств кибербезопасности (NCSC, CISA), провозглашающее принцип «Secure by Design» — безопасность должна быть встроенным свойством системы, а не надстройкой [60].

Заключение

Проведенное исследование позволяет сделать вывод о том, что безопасное использование персональных данных в системах искусственного интеллекта является сложной, многоаспектной проблемой, не имеющей единственного решения. Мы наблюдаем фундаментальный сдвиг парадигмы угроз: от защиты периметра к защите математической логики моделей и самих данных на протяжении всего их жизненного цикла.

Ключевые выводы:

- Недостаточность традиционного права: ни ФЗ-152, ни GDPR в их исходном виде не способны

полностью решить проблемы «черного ящика» и запоминания данных нейросетями. Механизмы экспериментальных правовых режимов (ЭПР) в РФ и «песочницы» в ЕС являются необходимым временным решением, но требуют трансформации в постоянные нормы, регулирующие использование синтетических данных и метрик приватности.

- Технологический императив: использование PETs (дифференциальная приватность, SMPC, HE) переходит из разряда академических экспериментов в категорию промышленных стандартов. Однако внедрение этих технологий требует существенных вычислительных мощностей и компетенций.
- MLSecOps как стандарт индустрии: безопасность ИИ должна стать неотъемлемой частью инженерной культуры. Внедрение практик MLSecOps и следование стандартам NIST/ISO является единственным способом минимизировать риски утечек и атак.

Перспективы:

Дальнейшие исследования должны быть направлены на разработку эффективных методов машинного разучивания (Unlearning), которые позволят реализовать «право на забвение» без полной перетренировки моделей, а также на создание гибридных схем шифрования, со-

четающих производительность SMPC и безопасность HE для сценариев реального времени. В условиях глобальной конкуренции в сфере ИИ, способность государства и бизнеса обеспечить баланс между инновационным развитием и защитой приватности станет ключевым фактором цифрового суверенитета.

ЛИТЕРАТУРА

1. A Survey on Privacy-Preserving Machine Learning Inference — Biblioteka Nauki.
2. (PDF) PRIVACY-PRESERVING MACHINE LEARNING: TECHNIQUES, CHALLENGES, AND FUTURE DIRECTIONS IN SAFEGUARDING PERSONAL DATA MANAGEMENT — ResearchGate.
3. Algorithms that remember: model inversion attacks and data protection law — PMC — NIH.
4. arXiv:2403.00871v1 [cs.CR] 1 Mar 2024.
5. ChatGPT Data Leaks and Security Incidents (2023-2025): A Comprehensive Overview, дата последнего обращения.
6. Whoops, Samsung workers accidentally leaked trade secrets via ChatGPT — Mashable.
7. Samsung's ChatGPT incident shows why AI needs a new security model — and why I think Filecoin might be it. — Reddit.
8. Статья 10. Специальные категории персональных данных — КонсультантПлюс.
9. The privacy-explainability trade-off: unraveling the impacts of differential privacy and federated learning on attribution methods — PubMed Central.
10. Zooming in on AI #15: Regulatory spaghetti and AI how to make sense of the EU GDPR and the EU AI Act — A&O Shearman.
11. Установлен экспериментальный правовой режим в сфере цифровых инноваций по направлению медицинской деятельности — КонсультантПлюс.
12. Кабмин утвердил экспериментально правовой режим для беспилотных авто в России.
13. The EU AI Act and the GDPR: collision or alignment? — Taylor Wessing, дата последнего обращения.
14. GDPR and AI Act: similarities and differences | activeMind.legal.
15. Digital Omnibus: EU Commission Proposes to Streamline GDPR and EU AI Act.
16. 2026: the year AI grows up.
17. Model inversion and membership inference: Understanding new AI security risks and mitigating vulnerabilities — Hogan Lovells.
18. How Model Inversion Attacks Compromise AI Systems — Defence.AI, дата обращения: 02 декабря 2025.
19. Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures — Thomas Ristenpart.
20. Model Inversion: The Essential Guide | Nightfall AI Security 101.
21. Membership Inference Attacks Against Machine Learning Models — Cornell: Computer Science.
22. A Method to Facilitate Membership Inference Attacks in Deep Learning Models, дата последнего обращения.
23. Range Membership Inference Attacks — arXiv.
24. Model Inversion Attacks: A Survey of Approaches and Countermeasures — arXiv.
25. Differential Privacy: Gradient Leakage Attacks in Federated Learning Environments — arXiv.
26. Gradient Leakage in Production Federated Learning — iQua Group.
27. Papers | Nicholas Carlini.
28. Extracting Training Data from Large Language Models.
29. Samsung employees allegedly leak data via ChatGPT — Cyber Security Hub.
30. Microsoft Actions Following Attack by Nation State Actor Midnight Blizzard, дата последнего обращения.
31. Update on Microsoft Actions Following Attack by Nation State Actor Midnight Blizzard.
32. 38TB of data accidentally exposed by Microsoft AI researchers | Wiz Blog.
33. Google Gemini — Wikipedia.
34. Gemini Data Breach: Analysis, Impact, and Lessons — LayerX Security.
35. Adversarial Misuse of Generative AI | Google Cloud Blog.
36. Differentially Private Finetuning for LLMs — Tom Ganz.
37. Differentially Private Fine-tuning of Language Models | Journal of Privacy and Confidentiality.
38. Evaluation of Open-Source Tools for Differential Privacy — MDPI.
39. HE-LRM: Encrypted Deep Learning Recommendation Models using Fully Homomorphic Encryption — arXiv.
40. Comparison of Fully Homomorphic Encryption and Garbled Circuit Techniques in Privacy-Preserving Machine Learning Inference — arXiv.
41. Recent advances of privacy-preserving machine learning based on (Fully) Homomorphic Encryption | Security and Safety (S&S).
42. Efficient Keyset Design for Neural Networks Using Homomorphic Encryption — MDPI.
43. CRYPTEN: Secure Multi-Party Computation Meets Machine Learning.
44. CrypTen Integration into PySyft — OpenMined.
45. [2109.00984] CrypTen: Secure Multi-Party Computation Meets Machine Learning — arXiv.

46. [Quick Review] CrypTen: Secure Multi-Party Computation Meets Machine Learning — Liner.
47. [2302.12580] Membership Inference Attacks against Synthetic Data through Overfitting Detection — arXiv.
48. Membership Inference Attacks against Synthetic Data through Overfitting Detection — Proceedings of Machine Learning Research.
49. Membership inference attacks against synthetic health data — PubMed — NIH.
50. The DCR Delusion: Measuring the Privacy Risk of Synthetic Data — arXiv.
51. MLSecOps: Protecting AI/ML Lifecycle in telecom — Ericsson.
52. The MLSecOps Blueprint for Securing the AI Development Lifecycle | by Tahir | Medium.
53. Building AI Security In: MLSecOps in Practice — YouTube.
54. A Survey of Machine Unlearning in Large Language Models: Methods, Challenges and Future Directions — arXiv.
55. AI Risk Management Framework | NIST — National Institute of Standards and Technology.
56. Artificial Intelligence Risk Management Framework (AI RMF 1.0) — NIST Technical Series Publications.
57. Secure Software Development Framework | CSRC.
58. ISO 42001 & NIST AI RMF: practical steps for mastering responsible AI governance in 2025.
59. Guidelines for secure AI system development.
60. Apple Workshop on Privacy-Preserving Machine Learning 2025.
61. Privacy enhancing technologies WHITE PAPER | Mastercard.
62. Homomorphic Encryption and Privacy Enhancing Technologies Webinar Series.
63. Privacy-Enhancing Technologies for Artificial Intelligence-Enabled Systems — arXiv.
64. Privacy-Enhancing Technologies in Federated Learning for the Internet of Healthcare Things: A Survey — MDPI.
65. Gemini Data Breach — Have I Been Pwned.
66. Model Inversion Attack: Unveiling Machine Learning Risks — Startup Defense.
67. Membership Inference Attacks Against Adversarially Robust Deep Learning Models — Princeton University.
68. Differential privacy — Wikipedia.
69. Understanding Differential Privacy | by An Nguyen | TDS Archive — Medium.
70. Shield Against Gradient Leakage Attacks: Adaptive Privacy-Preserving Federated Learning.
71. Protecting users with differentially private synthetic training data — Google Research.
72. Benchmarking Differentially Private Synthetic Data Generation Algorithms — Tumult Labs.
73. Empirical Analysis of Privacy-Fairness-Accuracy Trade-offs in Federated Learning: A Step Towards Responsible AI — arXiv.
74. A survey on machine unlearning: Techniques and new emerged privacy risks.
75. Data Reconstruction Attacks Against Machine Learning Models: Definition, Metrics, and Benchmark — arXiv.

© Воронин Сергей Анатольевич (info@voroninpravo.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

ОБ ОПРЕДЕЛЕНИИ ТРЕБОВАНИЙ ПО ЗАЩИТЕ ИНФОРМАЦИИ В ДОВЕРЕННЫХ ПРОГРАММНО-АППАРАТНЫХ КОМПЛЕКСАХ С УЧЕТОМ РЕЗУЛЬТАТИВНОСТИ МОНИТОРИНГА ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

ON THE DEFINITION OF INFORMATION SECURITY REQUIREMENTS IN TRUSTED SOFTWARE AND HARDWARE COMPLEXES, TAKING INTO ACCOUNT THE EFFECTIVENESS OF INFORMATION SECURITY MONITORING

A. Glukhov

Summary. The model and methodological approach to functional risk assessment and definition of system requirements for information protection in trusted software and hardware complexes (DPAC) for the purpose of their stable functioning are considered. A feature of the proposed approach is the ability to take into account reliability parameters DPAC, as well as the effectiveness of information security monitoring. The proposed approach can be applied to the creation and improvement of information security monitoring and management systems for trusted software and hardware complexes.

Keywords: trusted hardware and software complexes, reliability, protection from computer attacks, information security monitoring, functional risk.

Глухов Александр Александрович

АО «Научно-производственное объединение
«Критические информационные системы»

alexander.glukh0v@yandex.ru

Аннотация. Рассмотрены модель и методический подход к оцениванию функционального риска и определению системных требований по защите информации в доверенных программно-аппаратных комплексах (ДПАК) с целью их устойчивого функционирования. Особенностью предлагаемого подхода является возможность учета параметров надежности ДПАК, а также результативности мониторинга информационной безопасности. Предлагаемый подход может применяться при создании и совершенствовании систем мониторинга и управления информационной безопасностью доверенных программно-аппаратных комплексов.

Ключевые слова: доверенные программно-аппаратные комплексы, надежность, защищенность от компьютерных атак, мониторинг информационной безопасности, функциональный риск.

Федеральный закон от 26 июля 2017 года №187-ФЗ «О безопасности критической информационной инфраструктуры Российской Федерации» определяет безопасность объектов критической информационной инфраструктуры (КИИ) как состояние защищенности, обеспечивающее их устойчивое функционирование при компьютерных атаках. Устойчивое функционирование системы характеризуется способностью выполнять свои функции в условиях различных возмущающих факторов и восстанавливать их при нарушениях. Согласно национальному стандарту ГОСТ Р 57193-2025, устойчивость КИИ под угрозами достигается за счёт надежного функционирования системы с учетом требований по защите информации.

Особое внимание к устойчивости обусловлено возрастанием рисков информационной безопасности: атаки на инфраструктуру становятся всё более сложными, и полностью исключить инциденты невозможно. По-

этому в дополнение к традиционной информационной безопасности (ИБ) (нацеленной на защиту конфиденциальности, целостности и доступности данных) возникает концепция киберустойчивости, ставящая во главу угла непрерывность операций. На законодательном уровне закреплён принцип непрерывности и комплексности обеспечения безопасности КИИ, предполагающий, что меры защиты должны работать постоянно и во взаимосвязи, обеспечивая стойкость системы к сбоям. Практически это означает, что организации — субъекты КИИ — должны быть готовы не только отражать атаки, но и поддерживать ключевые сервисы в работе, быстро восстанавливаться после нарушений, минимизируя ущерб и перерывы в деятельности.

Определение системных требований по защите информации (ЗИ) в ДПАК — это установление правил или норм, которые должны соблюдаться при организации и осуществлении защиты информации, а также допустимых значений показателей эффективности защиты.

Технологическая независимость и безопасность КИИ подразумевают использование доверенных программно-аппаратных комплексов (ДПАК). При проведении замещения программно-аппаратных комплексов на объектах КИИ на отечественное оборудование основной проблемой в настоящее время является обеспечение требуемой надежности и производительности ДПАК, что необходимо учитывать при разработке моделей и методов определения требований по защите информации.

Для оценивания функциональных рисков информационной безопасности ДПАК [1] и определения системных требований по защите информации разработаны отдельные модели и методы [2–6].

Допустимые риски нарушения требований по ЗИ в процессе определения системных требований выступают в качестве характеристики количественных норм эффективности ЗИ и могут быть установлены в договорах, соглашениях и ТЗ с учетом специфики системы, а также могут быть определены, например, на основе прецедентного принципа или путем экспертного перевода качественных оценок допустимого риска в количественные значения. При этом экспертный метод преобразования качественных суждений о допустимом риске реализации угроз в количественные значения допустимой вероятности реализации угрозы основывается на теории нечетких суждений. Его применяют в том случае, когда статистика нарушений по возможным угрозам безопасности информации в рассматриваемой системе (и существующих аналогах системы) крайне мала, и не позволяет сформировать аргументированные исходные данные для моделирования, или отсутствует вообще.

Функциональный риск информационной безопасности ДПАК R_j может быть определен следующим образом:

$$R_j = \begin{cases} 0, & \text{при } q_{j\text{тр}} \leq q_j; \\ \frac{q_{j\text{тр}} - q_j}{q_{j\text{тр}} - q_{j\text{кр}}}; & \text{при } q_{j\text{кр}} < q_j < q_{j\text{тр}}; \\ 1; & \text{при } q_{j\text{кр}} \geq q_j \end{cases}$$

$q_{j\text{тр}}$ — требуемое значение j -го функционального показателя Пф, определяющее условие успешного (нормального) функционирования системы, при котором $R_j = 0$;

$q_{j\text{кр}}$ — характеризует предельно допустимое критическое значение деградации j -го функционального показателя Пф, при котором $R_j = 1$;

q_j — текущее значение j -го функционального показателя Пф в условиях компьютерных атак.

Для вероятностной модели на рисунке 1 приведен пример требуемого значения функционального показателя ДПАК ($q_{\text{тр}}$), его критического значения ($q_{\text{кр}}$) и поведения нижней границы траектории показателя (q) в условиях статистической неопределенности в параметрах воздействия КА.

Функциональный риск при этом определяется следующим образом:

$$R = \frac{q_{\text{тр}} - q}{q_{\text{тр}} - q_{\text{кр}}} \quad \text{при } q_{\text{кр}} < q_j < q_{\text{тр}}.$$

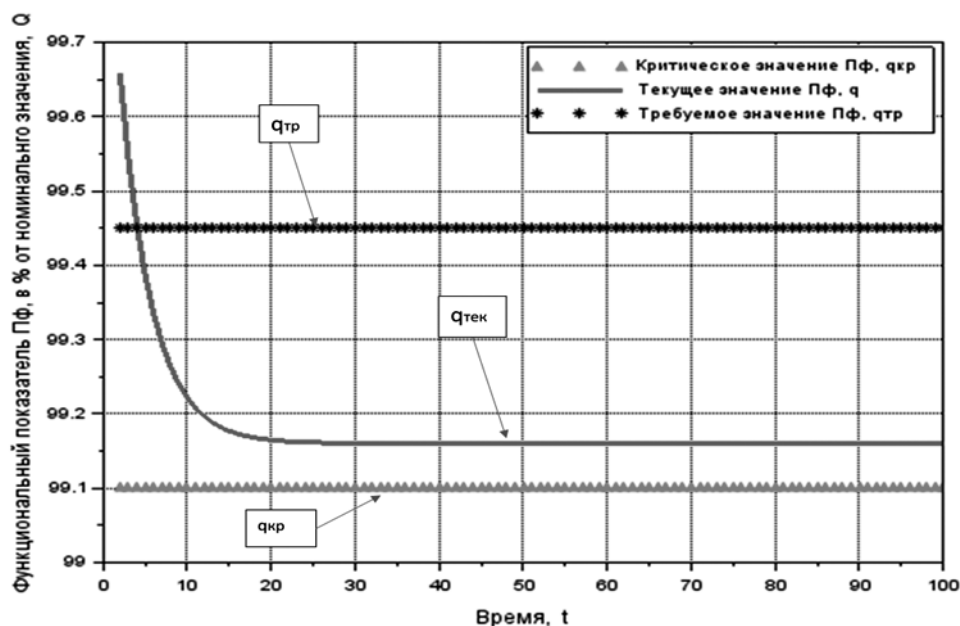


Рис. 1. Траектория функционального показателя Пф для определения рисков при наличии статистических данных (пример)

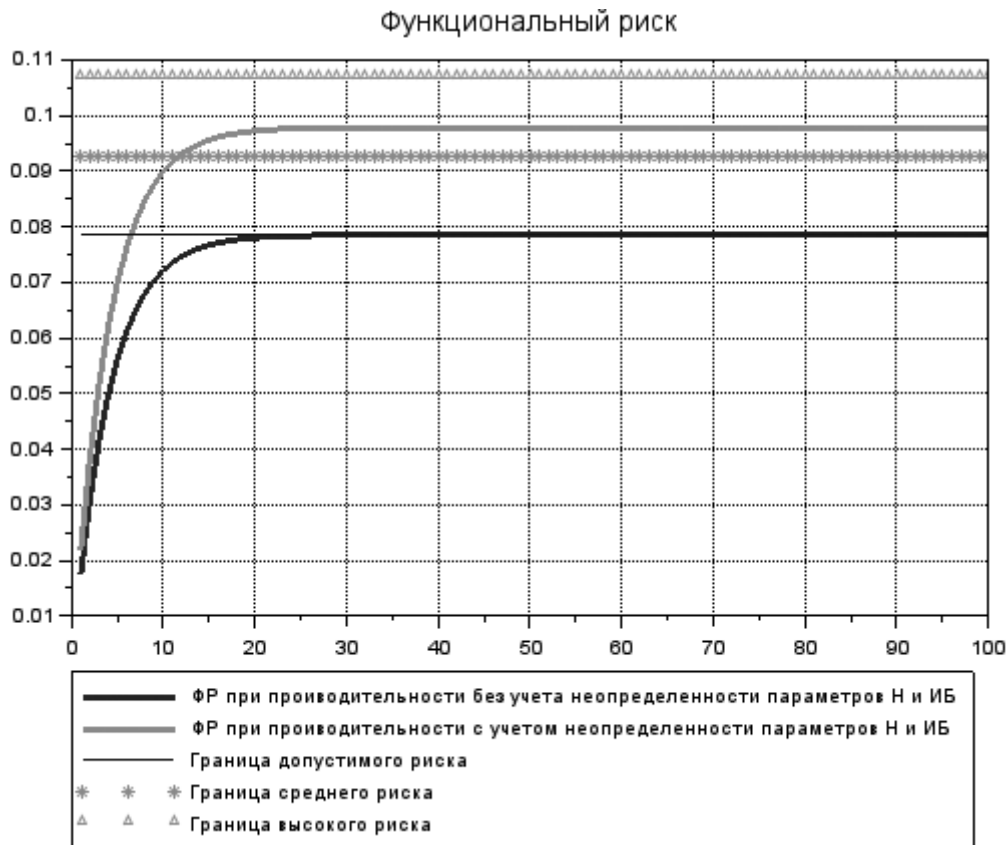


Рис. 2. Траектория и границы функционального риска при наличии статистических данных (пример)

Траектория и границы функционального риска при наличии статистических данных для данного случая представлен на рисунке 2.

При нечетких данных функциональный риск равен:

$$R = \frac{q_{тр} - q_{min}}{q_{тр} - q_{кр}} \quad \text{при} \quad q_{жкр} < q_j < q_{жтр};$$

где q_{min} — значение левой границы функции принадлежности (рисунок 3) показателя ПФ на определенном α -уровне (в данном случае равном 0,2).

Для оценивания функционального риска при нечетких данных необходимо определять степени принадлежности текущего значения функционального риска к термам из базы знаний. Они рассчитываются по-разному при четких и нечетких данных. При четких исходных данных степень принадлежности рассчитывается постановкой текущего значения переменной в формулу функции принадлежности. При нечетких исходных данных необходимо определить степень принадлежности одного нечеткого множества, соответствующего текущему значению риска утек к другим (эталонным) значениям, например, критическому $\mu_{кр}$, среднему $\mu_{ср}$, требуемому $\mu_{тр}$ (рисунок 4).

На основе вероятностных, нечетко-вероятностных и нечетких моделей можно решать различные задачи

оценки влияния компьютерных атак на бизнес-процессы, функциональные задачи в рамках этих бизнес-процессов и корпоративные информационные системы, обеспечивающих решение этих задач, а также оценки влияния на них неопределенностей в задании параметров моделей [2–6].

Рассмотрим модель оценивания требований по защите информации в ДПАК с учетом результативности системы мониторинга информационной безопасности.

Практически все системы мониторинга информационной безопасности (SIEM-системы) помогают реализовать ретроспективный анализ и организовать полноценное расследование в случае, если инцидент действительно случился. Принимаемые в настоящее время нормативные документы ФСБ России и ФСТЭК России как правило учитывают обязательное внедрение процессного подхода к построению систем защиты информации и мониторинга событий. Для выявления аномалий и угроз безопасности к SIEM-системе формируются следующие требования:

- корреляция событий на основе правил;
- корреляция событий на основе статистических методов.

Благодаря SIEM-системе, можно не просто выявить аномальные события, но и спрогнозировать их появление.

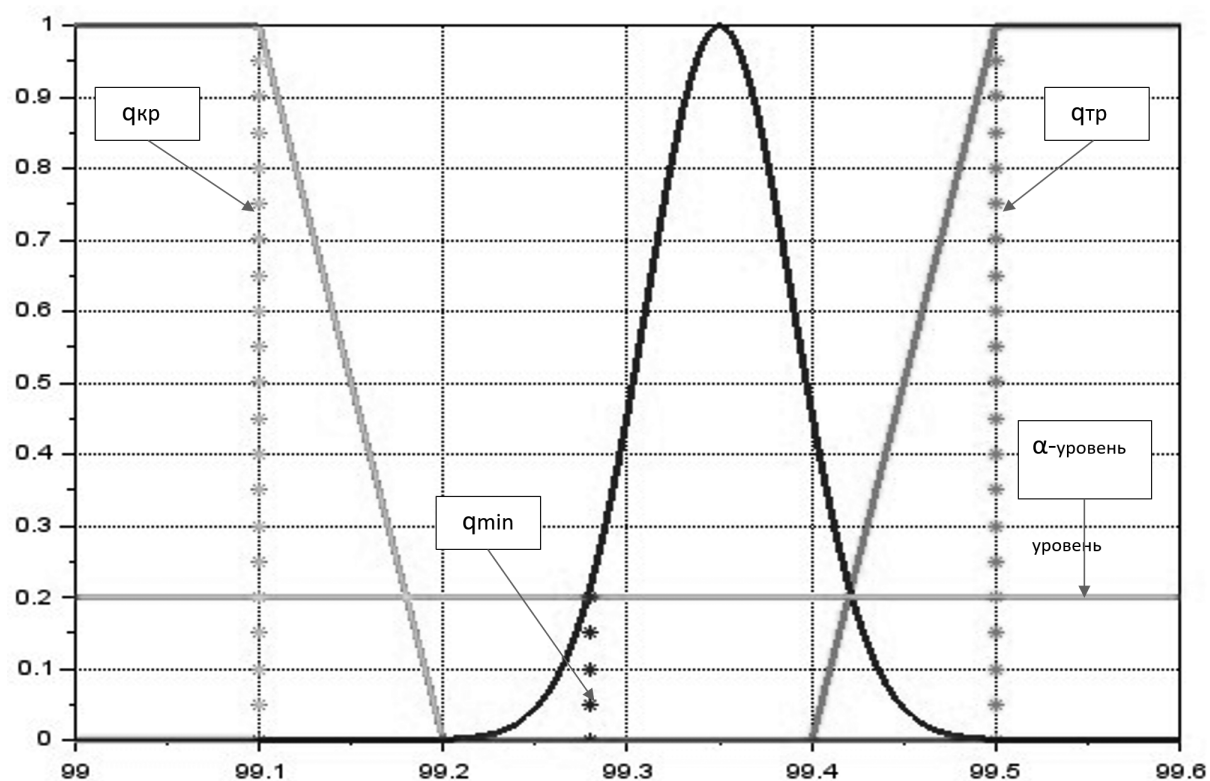


Рис. 3. Значения функций принадлежности показателей Пф для определения функциональных рисков при нечетких данных

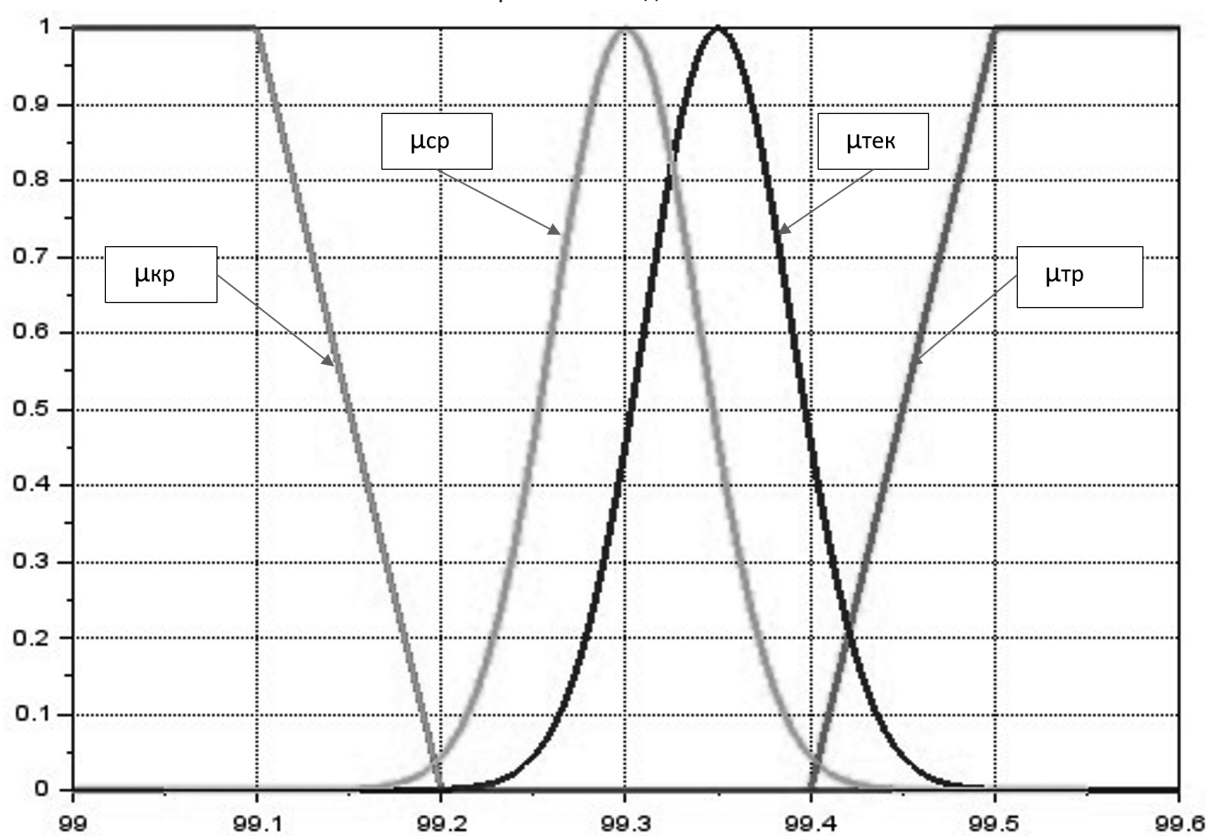


Рис. 4. Функции принадлежности для лингвистической переменной функционального показателя Пф при трехуровневом нечетком классификаторе

ние в будущем посредством машинного обучения, которое позволяет определять аномальные отклонения от нормы и ранжирует их по вероятности аномалии.

Возможности SIEM-систем могут быть применены для оценки рисков защищенности ДПАК, путем анализа вероятности возникновения и уровня критичности аномалии и инцидента информационной безопасности. Кроме того, SIEM-система может быть использована как основа системы интеллектуальных сервисов защиты информации, выработка управленческих решений в которой осуществляется путем обработки информации о событиях безопасности. К информации, характеризующей события безопасности, относятся все данные об изменении состояния элементов защищаемой инфраструктуры, формируемые программным или аппаратным способом, подлежащие хранению в электронном виде в специальных журналах в форме учетных записей (логов) или поступающие непосредственно в SIEM-систему по каналам связи. Однако, для методики оценки защищенности ДПАК возможно будет недостаточно стандартных правил обнаружения угроз.

Благодаря SIEM-системе, можно не просто выявить аномальные события, но и спрогнозировать их появление в будущем посредством машинного обучения, которое позволяет определять аномальные отклонения от нормы и ранжирует их по вероятности аномалии.

Рассмотрим модель ДПАК с проактивным мониторингом информационной безопасности (рисунок 5 а)), где

1 состояние — штатный режим функционирования ДПАК;

2 состояние — состояния восстановления работоспособности после прогнозирующего обнаружения КИ (предаварийный режим);

3 состояние — аварийный режим функционирования ДПАК в результате КА.

$(1-p_1)\lambda_{ка} = \lambda_{ки}$ — интенсивность КА, вызывающих компьютерные инциденты и нарушающих работоспособность ДПАК (аварийный режим);

$\lambda_{ка}$ — общее количество КА на ДПАК;

μ — интенсивность восстановления ДПАК из состояния, когда КИ вызвал аварийный режим функционирования ДПАК (например, «Отказ в обслуживании», далее отказ);

μ_2 — интенсивность восстановления ДПАК из предаварийного состояния;

p_1 — вероятность парирования КА средствами СЗИ;

p_2 — вероятность раннего (прогнозирующего) распознавания компьютерного инцидента на ранней стадии предаварийного состояния, когда нарушения функционирования ДПАК не происходит. Данный параметр характеризует эффективность системы мониторинга компьютерных атак.

Для случая проведения работ по восстановлению нормального функционирования ДПАК без нарушения ее работоспособности (к работоспособным относятся состояния 1 и 2) граф переходов представлен на рисунке 5 б).

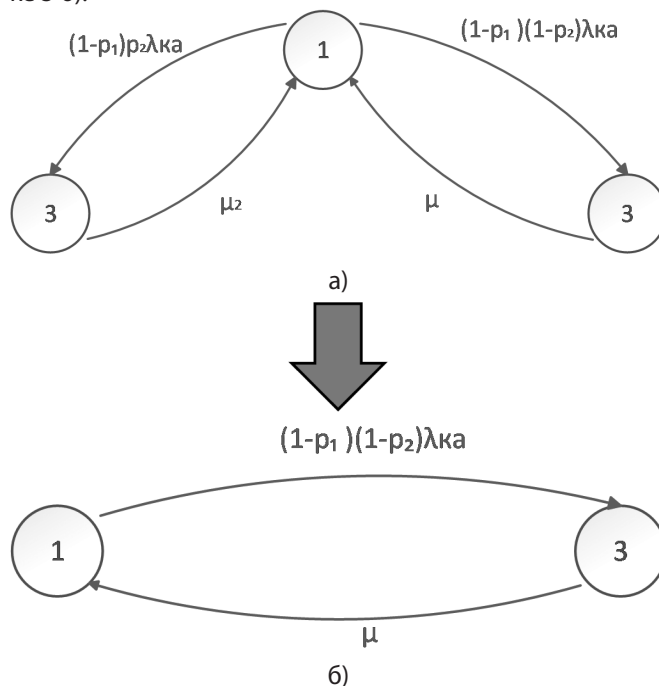


Рис. 5. Граф переходов модели ДПАК с прогнозирующим мониторингом КИ

Неотраженные СЗИ КА и необнаруженные на ранней стадии (восстановление происходит без остановки АСУ) КИ обнаруживаются с вероятностью p_3 системой мониторинга, далее ДПАК останавливается для восстановления (фактически нарушается функционирование ДПАК).

Граф переходов для данной модели ДПАК представлен на рисунке 6, где:

1 состояние — штатный режим функционирования ДПАК;

2 состояние — состояния восстановления функционирования после обнаружения предаварийного состояния (предотказа);

3 состояние — аварийный режим функционирования ДПАК в результате КА (отказ);

μ_1 — интенсивность восстановления ДПАК из аварийного состояния при его обнаружении системой мониторинга ИБ;

μ_2 — интенсивность восстановления ДПАК из аварийного состояния при его обнаружении штатными средствами (например, пользователями ДПАК) ($\mu_2 < \mu_1$).

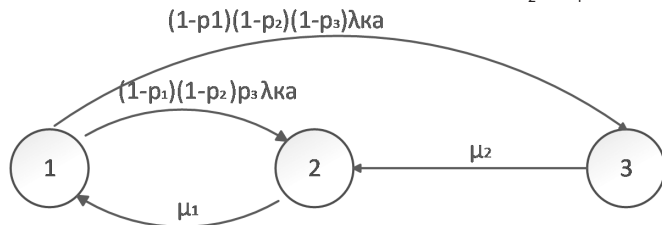


Рис. 6. Граф переходов модели с мониторингом КИ

Определим производительность ДПАК в виде (стационарный режим):

$$Q = Q_n \cdot K_{gn} \cdot K_{gka};$$

где

Q — фактическая производительность;

Q_n — номинальная производительность;

K_{gn} — коэффициент готовности, зависящий от факторов надежности;

$$K_{gn}(t) = \frac{\mu}{\lambda + \mu};$$

λ и μ — интенсивности отказов и восстановлений ДПАК;

и граф переходов для данной модели ДПАК представлен на рисунке 7, где:

1 состояние — штатный режим функционирования ДПАК;

2 состояние — состояния восстановления функционирования после отказа.

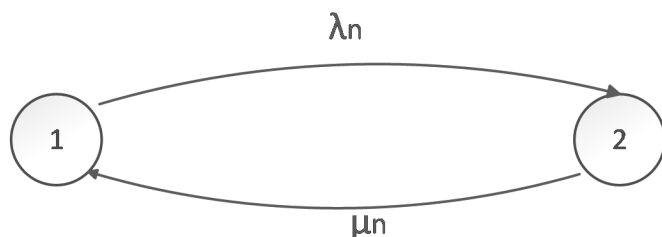


Рис. 7. Граф переходов модели ДПАК

K_{gka} — коэффициент готовности, зависящий от факторов защищенности ДПАК от компьютерных атак, в стационарном режиме.

Выражение для K_{gka} , полученное в результате решения системы уравнений, составленных для модели, представленной на рисунке 6, имеет вид:

$$K_{gka} = \frac{1}{\left(1 + \frac{(1-P_1)(1-P_2)P_3\lambda_{ka}}{\mu_1} + \frac{(1-P_1)(1-P_2)(1-P_3)\lambda_{ka}}{\mu_2}\right)}$$

В этом случае производительности ДПАК равна:

$$Q = Q_n \frac{\mu}{(\lambda + \mu)} \left(\frac{1}{\left(1 + \frac{(1-P_1)(1-P_2)P_3\lambda_{ka}}{\mu_1} + \frac{(1-P_1)(1-P_2)(1-P_3)\lambda_{ka}}{\mu_2}\right)} \right)$$

Функции влияния изменения параметров надежности и защищенности (интенсивностей отказов и восстановлений) на производительность ДПАК определяются следующим образом [5,6]:

$$Fw\lambda = \frac{dQ}{d\lambda} g_{\alpha} \sigma_{\lambda}; \quad Fw\mu = \frac{dQ}{d\mu} g_{\alpha} \sigma_{\mu};$$

$$Fw\lambda_{ka} = \frac{dQ}{d\lambda_{ka}} g_{\alpha} \sigma_{\lambda_{ka}}; \quad Fw\mu_1 = \frac{dQ}{d\mu_1} g_{\alpha} \sigma_{\mu_1};$$

$$Fw\mu_2 = \frac{dQ}{d\mu_2} g_{\alpha} \sigma_{\mu_2}; \quad Fwp1 = \frac{dQ}{dp1} g_{\alpha} \sigma_{p1};$$

$$Fwp2 = \frac{dQ}{dp2} g_{\alpha} \sigma_{p2}; \quad Fwp3 = \frac{dQ}{dp3} g_{\alpha} \sigma_{p3};$$

где

g_{α} — квантиль, соответствующая доверительной вероятности α ;

σ — среднеквадратические отклонения измерения параметров.

Выражения для частных производных производительности по параметрам надежности и защищенности могут быть получены с использованием математического пакета МАТКАД в результате символьных преобразований.

Частные функции влияния неопределенностей в определении параметров модели могут быть представлены в виде геометрической суммы

ошибок определения параметров надежности

$$FWSumNg = \sqrt{Fw\lambda^2 + Fw\mu^2};$$

ошибок определения параметров защищенности

Таблица 1.

Требуемые значения параметров защищенности и надежности ДПАК

	λ_{ka}	λ_n	μ_n	μ_1	μ_2	p_1	p_2	p_3
Начальные значения	0.00524	0.002	0.25	0.25	0.1	0.8	0.45	0.8
Требования при управлении по 8 параметрам	0.0044	0.0016	0.29	0.32	0.14	0.83	0.54	0.89
Требования при управлении по 7 параметрам		0.0016	0.29	0.34	0.16	0.84	0.57	0.91
Требования при управлении по 5 параметрам		0.0016	0.29			0.87	0.64	0.98

$$FWSumKAg =$$

$$= \sqrt{Fw\lambda_{ka}^2 + Fw\mu_1^2 + Fw\mu_2^2 + Fwp_1^2 + Fwp_2^2 + Fwp_3^2}.$$

Практическая реализация данного подхода к определению требований по защите информации вытекает, из того, что геометрическая сумма компонент некоторого вектора имеет минимальное значение в том случае, когда отдельные компоненты в геометрической сумме равны между собой, а их арифметическая сумма ограничена [5].

В этом случае величины требуемых изменений влияющих на производительность параметров защищенности (λ_{ka} , μ_1 , μ_2 , p_1 , p_2 , p_3) и надежности (λ_n , μ_n) для достижения производительности ДПАК не ниже требуемой

определяются выражениями (при возможности изменения всех 6-ти параметров защищенности):

$$d\lambda_n = \frac{FWSumNg / 2}{Fz\lambda_n}; \quad d\mu_n(t) = \frac{FWSumNg / 2}{Fz\mu}$$

$$d\lambda_{ka} = \frac{FWSumKAg / 6}{Fz\lambda_{ka}}; \quad d\mu_1(t) = \frac{FWSumKAg / 6}{Fz\mu_1};$$

$$d\mu_2(t) = \frac{FWSumKAg / 6}{Fz\mu_2}; \quad dp_1(t) = \frac{FWSumKAg / 6}{Fzp_1};$$

$$dp_2 = \frac{FWSumKAg / 6}{Fzp_2}; \quad dp_3 = \frac{FWSumKAg / 6}{Fzp_3}$$

В таблице 1 в качестве примера приведены результаты расчетов требуемых значений параметров защищенности, обеспечивающие заданный уровень функци-

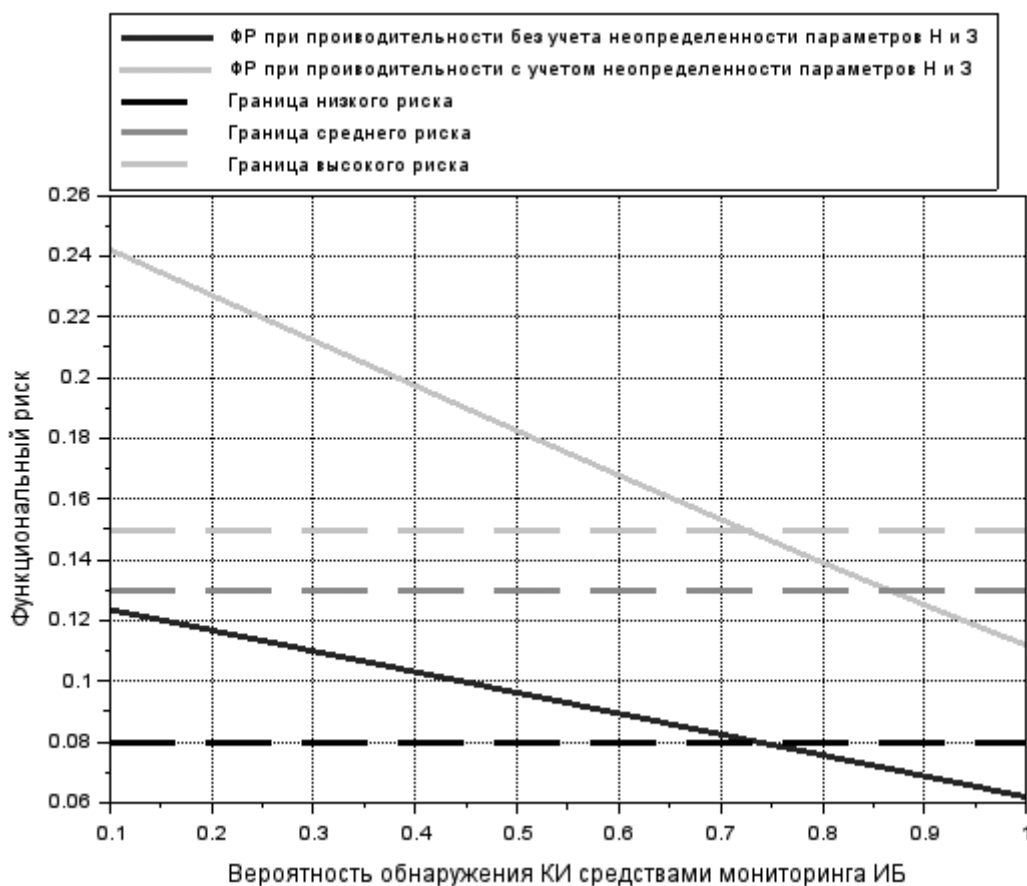


Рис. 7. Зависимость функционального риска ИБ от эффективности обнаружения компьютерных инцидентов

онального риска ИБ ДПАК. При расчетах принято: $P_1=0.8$, $P_2=0.45$, $P_3=0.8$. Минимально допустимое значение производительности ДПАК 99,6 % от номинального, при этом максимально допустимое значение функционального риска равно 0.08.

В данном случае при начальных значениях параметров защищенности с учетом их неопределенности функциональный риск ИБ ДПАК равен 0.14. При реализации требований по защите информации в ДПАК (таблица 1) функциональный риск ИБ равен 0.075, что меньше максимально допустимого.

Зависимость функционального риска ИБ от эффективности обнаружения компьютерных инцидентов средствами мониторинга ИБ приведена на рисунке 7.

Предложенная модель и подход к определению системных требований по защите информации в ДПАК

позволяет учитывать совместно такие факторы риска, как компьютерные атаки и факторы надежности, и решать определенные в нормативных документах (ГОСТы 59XXX— 2021г.) задачи:

- оценивание риска нарушения надежности функционирования системы в условиях отсутствия мер по ЗИ;
- оценивание риска нарушения функционирования системы в случае применения мер ЗИ, т. е. показатели остаточного риска при нарушении требований по ЗИ

без привязки к конкретным средствам защиты информации.

Реализация предложенного подхода позволяет также перейти к сбалансированному (рациональному) по критериям эффективность-стоимость построению систем защиты информации ДПАК.

ЛИТЕРАТУРА

1. Глухов А.П., Ададуров С.Е., Хмелевская Н.В., Глухов А.А. Оценка функциональности доверенных программно-аппаратных комплексов // Автоматика, связь, информатика. 2025. №3. С.12–15.
2. Глухов А.А., Глухов А.П., Корниенко А.А. О риск-ориентированном управлении параметрами защищенности доверенных программно-аппаратных комплексов // Защита информации. Инсайд. 2025. №2(122). С.52–57.
3. Глухов А.А., Белова Е.И., Глухов А.П. Управление параметрами надежности и защищенности доверенных программно-аппаратных комплексов // Современная наука: актуальные проблемы теории и практики: Серия «Естественные и Технические науки». 2024. №10. С.96–104.
4. Глухов А.А., Степанников Ю.М., Глухов А.П. Подход к управлению параметрами надежности и защищенности доверенных программно-аппаратных комплексов при нечетких данных // Современная наука: актуальные проблемы теории и практики: Серия «Естественные и Технические науки». 2025. №1. С.41–45.
5. Глухов А.П., Котяшев Н.Н., Лукин В.Л. Управление ресурсами проектируемых систем и комплексов критических приложений с заранее поставленными для них целями управления в условиях воздействий // Двойные технологии. 2008. №1. С. 46–55.
6. Василенко В.В., Котяшев Н.Н., Корнеев В.В. Аналитические представления процессов риска в комплексах и системах критических приложений // Двойные технологии. 2002. №1. С. 20–24.

© Глухов Александр Александрович (alexander.glukh0v@yandex.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»

ЭРГОНОМИЧЕСКАЯ ОЦЕНКА АВТО-ГЕНЕРАЦИИ UI С ИСПОЛЬЗОВАНИЕМ AI-АГЕНТА И MCP-СЕРВЕРА

ERGONOMIC EVALUATION OF UI AUTO-GENERATION USING AN AI AGENT AND AN MCP SERVER

**B. Goryachkin
G. Panov**

Summary. Problem statement. Automatic generation of UI components using AI agents and Model Context Protocol (MCP) is a promising approach; however, quantitative ergonomic data on this method are lacking.

Goal. To perform an ergonomic evaluation of AI-assisted code generation via MCP compared to manual development of UI components: to investigate time and quality characteristics.

Results. Based on six tasks, predictive linear models were built: MCP reduces operational load by 28, errors by 3, and execution time by 53–66 %.

Practical significance. The derived formulas enable pre-development effort estimation and identification of tasks where MCP is most beneficial.

Keywords: Model Context Protocol, ergonomics, AI agent, user interface components, UI-kit, Angular, linear regression.

Горячкин Борис Сергеевич

кандидат технических наук, доцент,
Московский государственный технический
университет им. Н.Э. Баумана
bsgor@mail.ru

Панов Герман Дмитриевич

Московский государственный технический
университет им. Н.Э. Баумана
german.panov.2018@mail.ru

Аннотация. Постановка проблемы. Автоматическая генерация UI-компонентов с использованием AI-агентов и Model Context Protocol (MCP) является перспективным направлением разработки пользовательских интерфейсов, однако количественные данные об эргономических характеристиках такого подхода отсутствуют.

Цель. Провести эргономическую оценку процесса автоматической генерации кода с использованием AI-агента и MCP: исследовать временные и качественные характеристики при создании UI-компонентов библиотеки по сравнению с ручным подходом.

Результаты. По результатам шести задач построены прогностические линейные модели: MCP снижает операционную нагрузку в 28 раз, количество ошибок — в 3 раза, время выполнения — на 53–66 %.

Практическая значимость. Полученные формулы позволяют планировать трудозатраты до начала разработки и определять задачи, для которых применение MCP наиболее эффективно.

Ключевые слова: Model Context Protocol, эргономика, AI-агент, компоненты пользовательского интерфейса, UI-kit, Angular, линейная регрессия.

Введение

Model Context Protocol (MCP) [1] — открытый протокол, разработанный Anthropic, который позволяет AI-агентам взаимодействовать с актуальной документацией в реальном времени [2]. Автоматическая генерация компонентов через MCP-сервер потенциально снижает время выполнения операций и количество ошибок. Ранее проведенные исследования инструментов AI-ассистированного программирования показали значительный прирост производительности, однако отсутствуют количественные данные об эргономических характеристиках работы с MCP.

Процесс исследования заключается в создании UI-компонентов с использованием UI-Kit для фреймворка Angular в ручном и автоматическом режимах.

Необходимо провести эргономическую оценку процесса автоматической генерации кода с использованием AI-агента и MCP: исследовать временные и каче-

ственные характеристики при создании UI-компонентов по сравнению с ручным подходом.

Задачи:

1. Выбрать критерии оценки для эксперимента (время, количество операций, ошибок, промптов).
2. Выполнить замеры на задачах разной сложности (простых, средних, сложных) для обоих подходов.
3. Рассчитать показатели улучшения эргономических характеристик, выявить зависимость эффективности от сложности задачи. Определить спектр задач, для которых применение MCP целесообразно.

Критерии оценки авто-генерации

Для количественной оценки эффективности двух подходов используются три критерия:

1. **Время выполнения** (t_{exec}) — полное время от начала задачи до получения работающего компонента, в минутах: $t_{\text{exec}} = \sum_{i=1}^n t_i$, где t_i — время i -го

этапа работы (поиск, написание кода, отладка и т. д.).

2. **Количество операций** (n_{ops}) — дискретные действия разработчика (клик, ввод, переход между

$$\text{окнами}): n_{ops} = \sum_{i=1}^k n_i$$

Для ручного подхода: n_{search} — поиск документации, n_{read} — чтение документации, n_{code} — написание кода, $n_{compile}$ — компиляция, n_{debug} — отладка.

Для MCP подхода: n_{prompt} — формулировка промптов, n_{review} — проверка результата, n_{adjust} — правки.

3. **Количество ошибок** (n_{errors}) — ошибки компиляции и runtime, требующие исправления:

$$n_{errors} = n_{syntax} + n_{import} + n_{type} + n_{runtime}$$

где: n_{syntax} — синтаксические ошибки, n_{import} — ошибки импортов, n_{type} — ошибки типизации, $n_{runtime}$ — ошибки времени выполнения.

Для выявления зависимостей применяется линейная регрессия вида $y = a \times C + b$, где C — коэффициент сложности задачи. Коэффициент детерминации R^2 показывает, насколько качество модели соответствует высокой предсказательной способности.

Коэффициент сложности задачи рассчитывается по формуле:

$$\begin{aligned} \text{Complexity} = \\ = w_1 \cdot C_{ui} + w_2 \cdot P + w_3 \cdot L + w_4 \cdot S + w_5 \cdot I \end{aligned} \quad (1)$$

где: C_{ui} — количество UI-компонентов библиотеки, действовавших в задаче; P — количество страниц (маршрутов) приложения; L — уровень сложности логики (валидация, CRUD, бизнес-правила); S — количество Angular-сервисов; I — количество внешних интеграций; $w_1 = 0.5, w_2 = w_4 = 2, w_3 = w_5 = 1$ — весовые коэффициенты, отражающие относительный вклад соответствующих параметров в сложность задачи.

На основе полученного коэффициента задачи распределяются по трем группам (табл. 1):

Таблица 1.

Группы задач

Группа	Диапазон	Характеристика
Группа А (простые)	до 8,0	Одна страница, 1–3 компонента, минимальная логика, без сервисов
Группа Б (средние)	8,1–20,0	Несколько страниц, 3–6 компонентов, валидация, роутинг, возможно 1 сервис
Группа В (сложные)	20,1+	Многостраничное приложение, 8+ компонентов, CRUD, несколько сервисов, интеграции

Экспериментальные исследования и оценка полученных результатов

Сравнительный эксперимент проводился на шести задачах разной сложности. Каждая задача выполнялась двумя способами:

Ручной подход: самостоятельный поиск документации, копирование примеров, отладка ошибок.

MCP подход: формулировка промптов для AI-агента (Claude Code [1]), который обращается к UI-Kit MCP Server и генерирует код на основе актуальной документации библиотеки.

Инструменты: Visual Studio Code, Node.js, UI-Kit, Angular. Начальное состояние для обоих подходов идентично: пустое Angular-приложение с установленными зависимостями UI-kit. Основу AI-агента составляет большая языковая модель класса Claude [4], дополненная контекстным доступом к документации через MCP.

Задачи эксперимента сгруппированы по сложности (табл. 2):

Таблица 2.

Разбиение задач по сложности по формуле (1)

Задача	C_{ui}	P	L	S	I	Итого	Группа
A1: Login	3	1	0	0	0	3,5	А (простая)
A2: Sign up	4	1	0	0	0	4,0	А (простая)
B1: Auth модуль	5	3	3	0	0	11,5	Б (средняя)
B2: Profile	6	2	3	0	3	13,0	Б (средняя)
B1: IoT Dashboard	10	4	6	3	6	31,0	В (сложная)
B2: Crypto Dashboard	12	5	6	4	6	36,0	В (сложная)

Каждая задача выполнялась двумя способами последовательно в одинаковых условиях. В табл. 3 и 4 зафиксированы: полное время выполнения, количество операций и ошибок, а также количество промптов для MCP-подхода. Сокращение времени рассчитывается как $\Delta t = (t_{manual} - t_{mcp}) / t_{manual}$.

Для анализа структуры операционной нагрузки каждая задача была разбита на отдельные типы действий. При ручном подходе доминируют три категории: написание кода (n_{code}), чтение (n_{read}) и поиск документации (n_{search}) — вместе они составляют около 80 % всех операций. MCP полностью устраняет поиск и чтение документации, заменяя их формулировкой промптов (n_{prompt}), количество которых растёт значительно медленнее.

Таблица 3.

Время и операции

Задача	C	$t_{\text{руч.}}$ МИН	$t_{\text{МСП}}$ МИН	ops _{руч.}	ops _{МСП}
A1: Login	3,5	35	12	60	3
A2: Sign up	4,0	45	18	70	4
B1: Settings	11,5	90	34	150	6
B2: Color Gen	13,0	105	41	170	7
B1: IoT Dash	31,0	300	138	400	15
B2: Crypto Dash	36,0	360	170	480	17

Таблица 4.

Ошибки, промпты и сокращение времени

Задача	C	Ошибки _{руч.}	Ошибки _{МСП}	Промпты	Δt
A1: Login	3,5	4	1	2	66 %
A2: Sign up	4,0	5	2	3	60 %
B1: Settings	11,5	8	3	4	62 %
B2: Color Gen	13,0	9	4	5	61 %
B1: IoT Dash	31,0	20	7	11	54 %
B2: Crypto Dash	36,0	24	8	13	53 %

Таблица 5.

Операции ручного подхода по категориям

Задача	C	n_{search}	n_{read}	n_{code}	$n_{\text{compilation}}$	n_{debug}	Итого
A1 (Login)	3,5	12	18	20	5	5	60
A2 (Sign up)	4,0	14	21	23	6	6	70
B1 (Settings)	11,5	30	40	50	15	15	150
B2 (Color Gen)	13,0	34	48	55	16	17	170

Задача	C	n_{search}	n_{read}	n_{code}	$n_{\text{compilation}}$	n_{debug}	Итого
B1 (IoT Dash)	31,0	80	120	130	35	35	400
B2 (Crypto Dash)	36,0	96	144	156	42	42	480

Таблица 6.

Операции МСП-подхода по категориям

Задача	C	n_{prompt}	n_{review}	n_{adjust}	Итого
A1 (Login)	3,5	2	1	0	3
A2 (Sign up)	4,0	3	1	0	4
B1 (Settings)	11,5	4	1	1	6
B2 (Color Gen)	13,0	5	1	1	7
B1 (IoT Dash)	31,0	11	2	2	15
B2 (Crypto Dash)	36,0	13	2	2	17

Суммарный коэффициент роста операций: ручной — 12,76; МСП — 0,45 (в 28 раз меньше). На рисунке 1 наглядно видно, как с ростом сложности задачи нагрузка при ручном подходе нарастает многократно быстрее, тогда как при МСП остаётся минимальной.

Анализ ошибок проводился по четырём категориям для каждого подхода (табл. 7).

По результатам линейной регрессии для каждой подкатегории операций и ошибок построены прогностические модели (табл. 8) вида $y = a \times C + b$, где C — комплексный показатель сложности задачи. Коэффициент a характеризует прирост показателя при увеличении сложности на 1 единицу, а b отражает базовый уровень для простейшей задачи.

На рисунке 3 представлены зависимости времени выполнения и числа промптов от сложности задачи. График (а) показывает, что время при ручном подходе

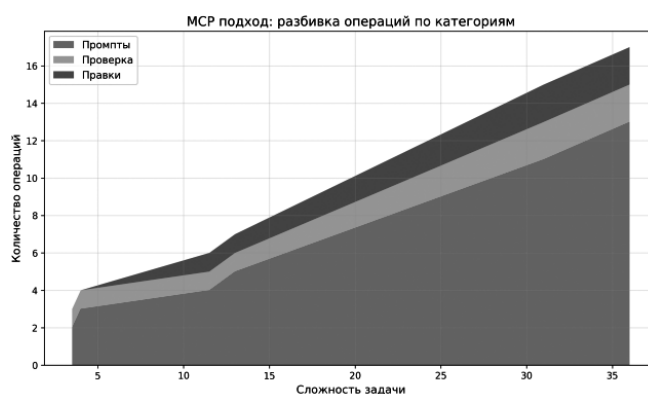
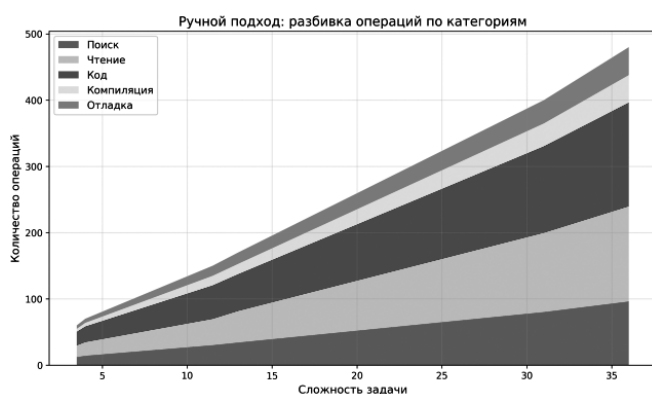


Рис. 1. Разбивка операций по категориям: ручной и МСП подходы

Таблица 7.

Ошибки по типам: синтаксические, импорты, типизация, runtime

3-10 Задача	C	Ручной подход				МСП подход			
		n_{syn}	n_{imp}	n_{type}	n_{rt}	n_{syn}	n_{imp}	n_{type}	n_{rt}
A1 (Login)	3,5	0	2	1	1	0	0	1	0
A2 (Sign up)	4,0	1	2	1	1	0	0	1	1
B1 (Settings)	11,5	1	3	2	2	0	1	1	1
B2 (Color Gen)	13,0	1	3	3	2	0	1	2	1
B1 (IoT Dash)	31,0	2	8	5	5	0	2	3	2
B2 (Crypto Dash)	36,0	3	9	6	6	0	2	4	2

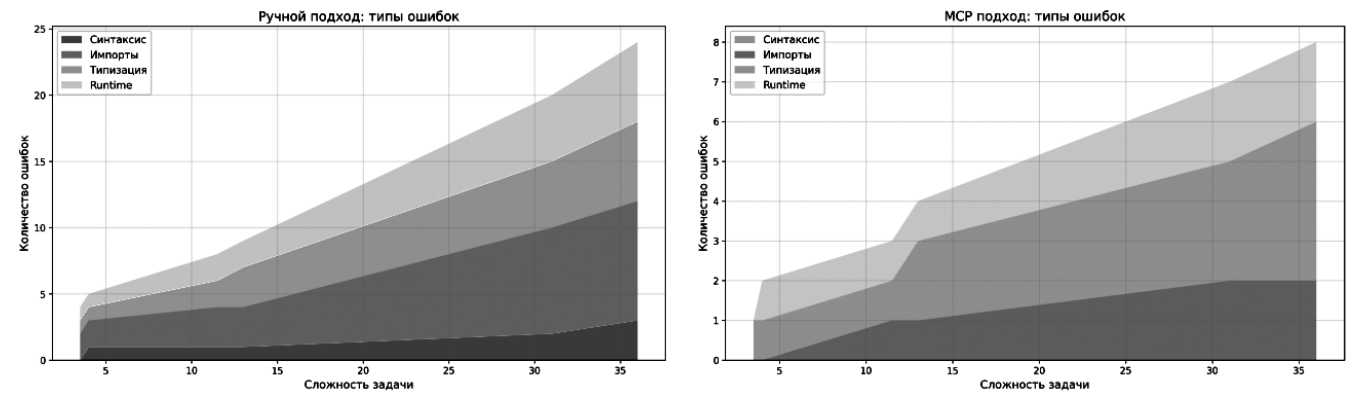


Рис. 2. Структура ошибок по типам: ручной и МСП подходы

Таблица 8.
Сводная таблица прогностических формул

Категория	Формула	R^2
ОПЕРАЦИИ — Ручной подход		
Чтение документации	$n_{read} \approx 3,87 \times C + 1,39$	0,994
Написание кода	$n_{code} \approx 4,13 \times C + 4,17$	0,998
Поиск документации	$n_{search} \approx 2,55 \times C + 2,23$	0,998
Компиляция	$n_{compile} \approx 1,11 \times C + 1,56$	0,999
Отладка	$n_{debug} \approx 1,10 \times C + 1,79$	0,998
Суммарно	$n_{ops} \approx 12,76 \times C + 11,14$	0,998
ОПЕРАЦИИ — МСП подход		
Промпты	$n_{prompt} \approx 0,35 \times C + 0,93$	0,989
Проверка	$n_{review} \approx 0,04 \times C + 0,74$	0,910
Правки	$n_{adjust} \approx 0,06 \times C - 0,03$	0,929
Суммарно	$n_{ops} \approx 0,45 \times C + 1,64$	0,985

Категория	Формула	R^2
ОШИБКИ — Ручной подход		
Импорты (38 % всех)	$n_{import} \approx 0,23 \times C + 0,78$	0,980
Типизация	$n_{type} \approx 0,15 \times C + 0,52$	0,982
Runtime	$n_{runtime} \approx 0,15 \times C + 0,29$	0,993
Синтаксис	$n_{syntax} \approx 0,07 \times C + 0,18$	0,870
Суммарно	$n_{errors} \approx 0,60 \times C + 1,77$	0,982
ОШИБКИ — МСП подход		
Типизация	$n_{type} \approx 0,09 \times C + 0,55$	0,926
Импорты	$n_{import} \approx 0,06 \times C - 0,03$	0,929
Runtime	$n_{runtime} \approx 0,05 \times C + 0,35$	0,818
Синтаксис	$n_{syntax} = 0$	1,000
Суммарно	$n_{errors} \approx 0,20 \times C + 0,87$	0,921

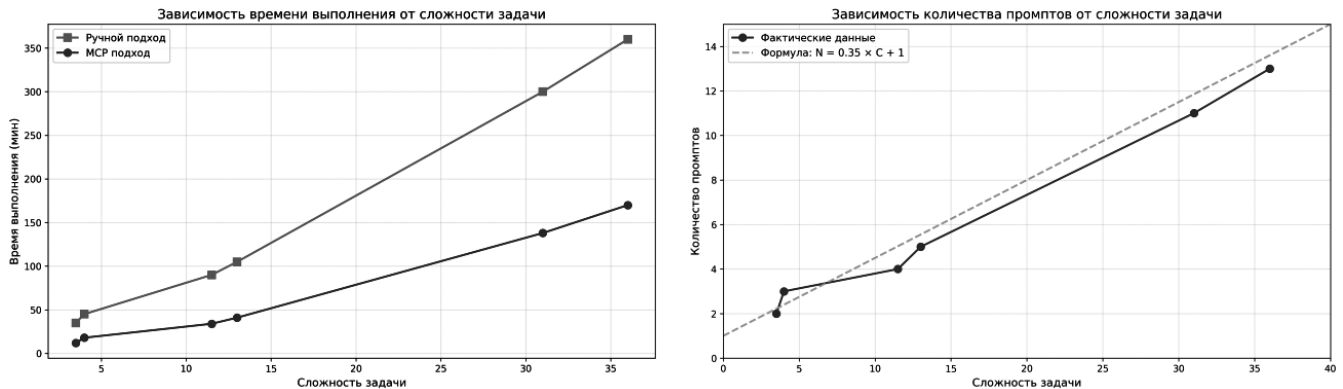


Рис. 3. Зависимость времени выполнения и числа промптов от сложности задачи

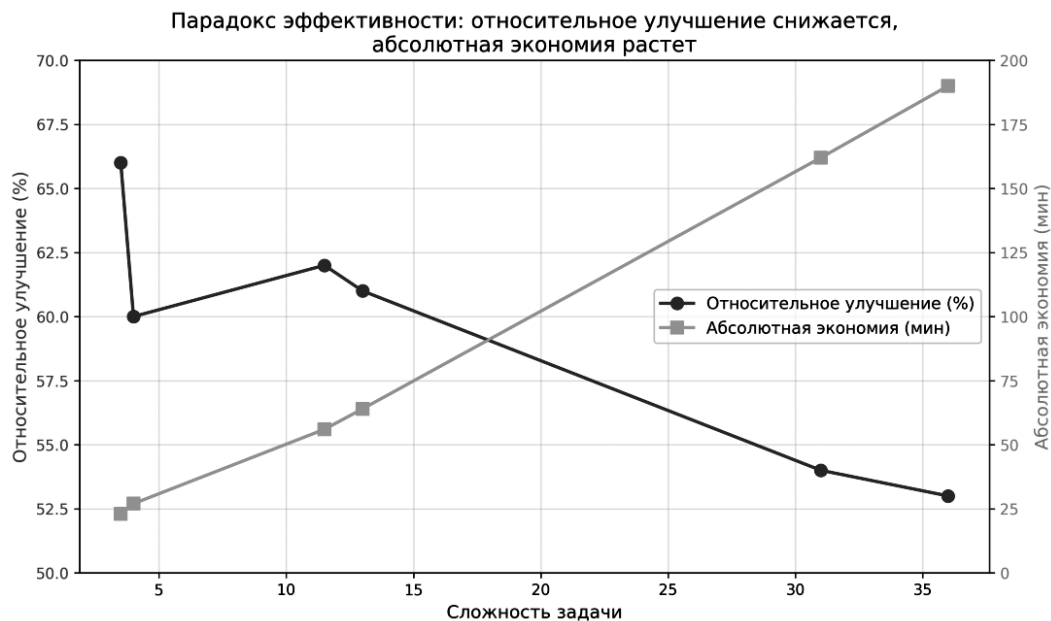


Рис. 4. Парадокс эффективности: относительное улучшение снижается, абсолютная экономия растёт

растёт значительно быстрее с ростом сложности. График (б) подтверждает линейный характер роста числа промптов, что позволяет заранее планировать итерации взаимодействия с агентом.

На рисунке 4 проиллюстрирован «парадокс эффективности»: относительный выигрыш ($\Delta t\%$) снижается от 66 % до 53 % при росте сложности, тогда как абсолютная экономия в минутах нарастает — с 23 до 190 минут. Это объясняется тем, что базовое время ручного подхода растёт быстрее, чем линейный прирост числа промптов при МСП.

Исследование выявило устойчивые линейные зависимости между сложностью задачи и всеми эргономическими критериями. МСП полностью устраняет операции поиска и чтения документации, что математически выражается в коэффициенте n_{ops} , сниженном в 28 раз: с 12,76 до 0,45. Ошибки типизации остаются наименее поддающейся автоматизации категорией, что согласуется с результатами работы [5], где аналогичные ограничения вы-

явлены для GitHub Copilot. Это требует дополнительной верификации сгенерированного кода.

Выявлен «парадокс эффективности»: относительное улучшение времени снижается (66 % → 53 %) при росте сложности, тогда как абсолютная экономия увеличивается (23 → 190 мин). Причина — линейный рост числа промптов ($N \approx 0,35 \times C + 1$) при значительно более быстром росте времени ручного подхода. По наблюдениям [3], итеративное уточнение промптов является ключевой составляющей эффективной работы с AI-инструментами, что подтверждается полученной моделью роста n_{prompt} .

Пример практического применения моделей. Предположим, разработчику предстоит реализовать форму фильтрации с 3 полями ввода, выпадающим списком и кнопкой сброса. По формуле 1 ($C_{ui} = 3, P = 1, L = 0, S = 0, I = 0$):

$$C = 0,5 \times 3 + 2,0 \times 1 + 0 + 0 + 0 = 3,5$$

Подставляя $C = 3,5$ в модели из табл. 8, получаем оценки до начала работы (табл. 9):

Таблица 9.

Оценки на основе прогностических формул

Показатель	Ручной	МСП
Операций	$12,76 \times 3,5 + 11,14 \approx 56$	$0,45 \times 3,5 + 1,64 \approx 3$
Ошибок	$0,60 \times 3,5 + 1,77 \approx 4$	$0,20 \times 3,5 + 0,87 \approx 2$
Промптов	$0,35 \times 3,5 + 1 \approx 2$	

МСП снижает число операций в 18 раз ($56 \rightarrow 3$) и число ошибок вдвое ($4 \rightarrow 2$). Прогнозируемое число промптов (2) соответствует 1–2 итерациям уточнения.

Операционная нагрузка: ручной — $n_{ops} \approx 12,76C + 11,14$; МСП — $n_{ops} \approx 0,45C + 1,64$ (снижение в 28 раз)

Ошибки: ручной — $n_{errors} \approx 0,60C + 1,77$; МСП — $n_{errors} \approx 0,20C + 0,87$ (снижение в 3 раза)

Промпты: $N \approx 0,35C + 1$; позволяет планировать итерации взаимодействия с агентом

Заключение

Проведена эргономическая оценка автоматической генерации кода с использованием AI-агента и МСП на примере UI-Kit. По результатам шести задач построены прогностические линейные модели.

Время выполнения сокращается на 53–66 %; операционная нагрузка — на 94–96 %; количество ошибок — на 60–75 %. МСП оптимален для задач, где абсолютная экономия превышает 1 час.

Направления дальнейших исследований: расширение диапазона сложности, привлечение нескольких испытуемых, тестирование на других библиотеках компонентов, оценка долгосрочного влияния на качество кодовой базы.

ЛИТЕРАТУРА

1. Anthropic. Claude code documentation: MCP servers. <https://docs.anthropic.com/en/docs/claude-code>, 2026. Официальная документация Claude Code — интерактивного инструмента командной строки для разработки с использованием AI-агентов.
2. Eirini Kalliamvakou. Research: quantifying GitHub Copilot's impact on developer productivity and happiness. GitHub Blog, 2022.
3. Christian Bird, Denae Ford, Thomas Zimmermann, Nicole Forsgren, Eirini Kalliamvakou, Travis Lowdermilk, and Idan Gazit. Taking flight with Copilot: Early insights and opportunities of AI-powered pair-programming tools. Queue, 20(6):35–57, 2023.
4. Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harrison Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, et al. Evaluating large language models trained on code. arXiv preprint arXiv:2107.03374, 2021.
5. Arghavan Moradi Dakhel, Vahid Majdinasab, Amin Nikanjam, Foutse Khomh, Michel C. Desmarais, and Zhen Ming Jiang. GitHub Copilot AI pair programmer: Asset or Liability? Journal of Systems and Software, 203:111734, 2023.
6. Google. Angular documentation. <https://angular.io/docs>, 2026.
7. Xinyi Hou, Yanjie Zhao, Shenao Wang, and Haoyu Wang. Model context protocol (MCP): Landscape, security threats, and future research directions. arXiv preprint arXiv:2503.23278, 2025.
8. Anthropic. Model context protocol documentation. <https://modelcontextprotocol.io/>, 2026. Официальная документация Model Context Protocol.
9. ГОСТ Р ИСО 9241-11-2010. Эргономика взаимодействия человек-система.
10. Часть 11. Руководство по пригодности использования. Росстандарт, 2010.
11. Б.С. Горячкин. Эргономический анализ систем обработки информации и управления. МГТУ им. Н.Э. Баумана, 2015.

© Горячкин Борис Сергеевич (bsgor@mail.ru); Панов Герман Дмитриевич (german.panov.2018@mail.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

АНАЛИЗ ВРЕМЕНИ ОТКЛИКА НА ПОИСКОВЫЕ ЗАПРОСЫ В СИСТЕМАХ ПОИСКА НАУЧНЫХ ПУБЛИКАЦИЙ

ANALYSIS OF QUERY RESPONSE TIME IN SCIENTIFIC PUBLICATION RETRIEVAL SYSTEMS

**B. Goryachkin
P. Sysa**

Summary. Problem statement. In modern systems for searching scientific publications, response time is a key parameter determining the effectiveness of user interaction with the system. The growth of data volumes and the increasing complexity of query processing algorithms lead to higher computational costs and delays in generating search results, which necessitates a formal analysis of the time cost structure.

Goal. Development of an analytical model of response time in scientific publication search systems and analysis of the contribution of individual query processing stages to the overall response time when using different search approaches.

Results. A mathematical model of response time is developed as a sum of time costs for query preprocessing, query interpretation, candidate document set generation, relevance computation, ranking, and response generation for the user. An interpretation of the model coefficients reflecting computational costs and characteristics of the hardware and software environment is provided. The dependence of response time on the applied search approaches — keyword, semantic, and hybrid — is analyzed.

Practical significance. The proposed model enables quantitative evaluation of the performance of scientific publication search systems, identification of the most computationally expensive stages of query processing, and justification of algorithmic and architectural choices under response time constraints.

Keywords: information retrieval, response time, search engines, query processing, mathematical modeling, performance, semantic search, hybrid search.

Горячкин Борис Сергеевич

кандидат технических наук, доцент,
Московский государственный технический
университет им. Н.Э. Баумана
bsgor@mail.ru

Сыса Павел Валерьевич

Московский государственный технический
университет им. Н.Э. Баумана
sysa.pavel@yandex.ru

Аннотация. Постановка проблемы. В современных системах поиска научных публикаций время отклика является ключевым параметром, определяющим эффективность взаимодействия пользователя с системой. Рост объёма данных и усложнение алгоритмов обработки запросов приводят к увеличению вычислительных затрат и задержек при формировании поисковой выдачи, что требует формализованного анализа структуры временных затрат.

Цель. Разработка аналитической модели времени отклика системы поиска научных публикаций и исследование вклада отдельных этапов обработки запроса в общее время отклика при использовании различных поисковых подходов.

Результаты. Разработана математическая модель времени отклика в виде суммы временных затрат на этапы предварительной обработки запроса, интерпретации запроса, формирования кандидатного множества документов, вычисления релевантности, ранжирования и формирования ответа пользователю. Выполнена интерпретация коэффициентов модели, отражающих вычислительные затраты и особенности аппаратно-программной среды. Проведён анализ зависимости времени отклика от используемых поисковых подходов: ключевого, семантического и гибридного.

Практическая значимость. Предложенная модель позволяет количественно оценивать производительность систем поиска научных публикаций, выявлять наиболее вычислительно затратные этапы обработки запросов и обосновывать выбор алгоритмов и архитектурных решений с учётом требований к времени отклика.

Ключевые слова: информационный поиск, время отклика, поисковые системы, обработка запроса, математическое моделирование, производительность, семантический поиск, гибридный поиск.

Введение

Современные системы поиска научных публикаций функционируют в условиях постоянного роста объёма научной информации, что приводит к повышенным требованиям к времени отклика и скорости обработки пользовательских запросов. Одним из ключевых показателей качества работы поисковой системы является время отклика, напрямую влияющее на эффективность работы пользователя и удобство взаимодействия с системой.

Увеличение времени отклика приводит к росту когнитивной нагрузки и снижению эффективности взаимодействия пользователя с системой. В связи с этим возникает необходимость формализации процесса обработки поискового запроса и аналитической оценки временных затрат на его основных этапах.

Целью работы является разработка аналитической модели времени отклика системы поиска научных публикаций и анализ временных характеристик обработки поисковых запросов с учётом архитектурных и алгорит-

мических особенностей, что позволяет сопоставить различные поисковые подходы (ключевой, семантический, гибридный) с точки зрения временных затрат.

Поисковые подходы нахождения научных публикаций

Рассматриваются три поисковых подхода: ключевой, семантический и гибридный.

Ключевой поиск основан на сопоставлении терминов пользовательского запроса с терминами документов. Документы предварительно индексируются с использованием инвертированной структуры, в которой каждому термину соответствует список содержащих его документов. Обработка запроса включает лексическую нормализацию, обращение к словарю индекса, формирование множества кандидатных документов на основе операций над постинговыми списками и последующее ранжирование с применением статистических частотных моделей. Данный подход основывается на прямом совпадении терминов и использует преимущественно лексические признаки [1].

Семантический поиск основан на применении распределённых векторных представлений текста. Документы и запрос отображаются в латентное пространство фиксированной размерности. Множество кандидатных документов формируется посредством вычисления меры сходства между вектором запроса и векторами документов с использованием специализированных структур поиска ближайших соседей. В отличие от ключевого подхода, сопоставление осуществляется на уровне векторных представлений, что позволяет учитывать контекстуальные и синонимические зависимости [6, 8].

Гибридный подход сочетает ключевой отбор с последующим семантическим уточнением. Сначала формируется ограниченное множество кандидатных документов с использованием инвертированного индекса, затем выполняется их семантическое переранжирование. Такой способ организации обработки запроса позволяет объединить эффективность ключевого поиска и преимущества векторного сопоставления.

Общая модель времени отклика

Приняты следующие допущения: обработка запроса представляется как последовательность основных этапов. Корпус документов считается предварительно проиндексированным, затраты на построение и обновление индекса не учитываются. Аппаратная платформа считается однородной, коэффициенты временных затрат — усреднёнными. Основной вклад в общее время отклика вносят вычислительные операции обработки запроса и доступа к индексным структурам, сетевые

задержки не учитываются. Формирование поисковой выдачи выполняется при ограниченном размере результирующего списка документов с использованием частичного ранжирования.

Под временем отклика понимается интервал времени от момента отправки пользователем поискового запроса до получения отсортированной поисковой выдачи [2]. Время отклика системы поиска научных публикаций представляется в виде суммы времён выполнения основных этапов обработки запроса:

$$T_{search} = T_{prep} + T_{interp} + T_{cand} + T_{rel} + T_{rank} + T_{resp} \quad (1)$$

где:

- T_{search} — общее время отклика системы поиска.
- T_{prep} — время предварительной обработки пользовательского запроса.
- T_{interp} — время интерпретации поискового запроса.
- T_{cand} — время формирования множества кандидатных документов.
- T_{rel} — время вычисления релевантности документов.
- T_{rank} — время ранжирования документов.
- T_{resp} — время формирования и выдачи ответа пользователю.

Каждое слагаемое соответствует отдельному функциональному этапу обработки запроса, выполнение которых происходит последовательно, что отражает конвейерный характер обработки и позволяет декомпозировать общее время отклика на отдельные компоненты для оценки эффективности оптимизаций, направленных на конкретные этапы поискового процесса.

Декомпозиция T_{search} соответствует модели информационного поиска, в которой обработка запроса рассматривается как последовательность этапов: предварительная обработка, интерпретация, обращение к индексным структурам, вычисление оценок релевантности и формирование ранжированной выдачи [1]. Аддитивная форма модели следует из принципа суммирования времен последовательных алгоритмических стадий [12]. Индексация не включается в модель, так как выполняется вне интерактивного контура, и в расчет принимаются только операции, осуществляемые после отправки запроса, что соответствует определению времени отклика как интервала между действием пользователя и получением результата [2].

Этап предварительной обработки предназначен для приведения пользовательского запроса к унифицированному виду, используемому системой поиска, и включает операции токенизации, нормализации

и лемматизации [1]. Данный этап играет ключевую роль в обеспечении корректности последующих вычислений, поскольку устраняет лингвистические вариации и неоднозначности, связанные с формой записи терминов, а также влияет на полноту и точность поиска и стабильность временных характеристик системы.

$$T_{prep} = C_{prep} \times L_q \quad (2)$$

где:

- C_{prep} — среднее время предварительной обработки одного термина.
- L_q — количество терминов в пользовательском запросе.

Линейный характер зависимости обусловлен тем, что операции выполняются независимо для каждого термина. Значение коэффициента C_{prep} определяется сложностью используемых лингвистических алгоритмов, языком запроса и особенностями аппаратной реализации, включая SIMD-инструкции, векторизацию и многопоточность. Увеличение длины запроса приводит к пропорциональному росту времени выполнения данного этапа.

Этап интерпретации запроса обеспечивает сопоставление пользовательских терминов с внутренними представлениями системы, такими как словари, тезаурусы и семантические модели, и реализует переход от текстового представления запроса к формализованным структурам, используемым системой поиска. Данный этап особенно важен для обеспечения корректного сопоставления пользовательских формулировок с документами корпуса.

$$T_{interp} = L_q \times (C_{dict} \times \log_2 V + C_{sem}) \quad (3)$$

где:

- L_q — количество терминов в пользовательском запросе.
- C_{dict} — среднее время поиска одного термина в словаре.
- V — размер словаря или тезауруса системы.
- C_{sem} — среднее время семантической интерпретации одного термина.

Логарифмическая зависимость обусловлена логарифмической сложностью поиска (например, бинарный поиск и В-деревья) [12]. Основание логарифма выбрано равным 2, что соответствует числу шагов бинарного поиска, при этом выбор основания не влияет на асимптотическое поведение модели, так как учитывается коэффициентом C_{dict} . Семантическая интерпретация учитывается отдельным слагаемым, поскольку её вычислительная сложность не зависит от размера словаря.

Этап формирования кандидатного множества предназначен для извлечения документов, потенциально релевантных пользовательскому запросу, из индексных структур [1], и определяет масштаб последующих вычислений, поскольку размер кандидатного множества напрямую влияет на время вычисления релевантности и ранжирования документов.

$$T_{cand} = \sum_{i=1}^{L_q} (C_{read} \times F_i) + C_{merge} \times N_c \quad (4)$$

где:

- L_q — количество терминов в пользовательском запросе.
- F_i — количество документов, содержащих i -й термин запроса.
- C_{read} — среднее время чтения одного элемента индексного списка.
- C_{merge} — среднее время обработки одного документа при объединении индексных списков.
- N_c — количество документов в кандидатном множестве.

Первая часть формулы отражает чтение постинговых списков для каждого термина запроса, вторая часть — операции объединения, фильтрации и устранения дубликатов документов. Частотные термины приводят к росту значений F_i , что увеличивает время выполнения этапа и может оказывать доминирующее влияние на общее время отклика [1]. Формирование множества кандидатных документов рассматривается в унифицированном виде для всех поисковых подходов: предполагается предварительный отбор ограниченного множества документов-кандидатов размером N_c , а для семантического поиска конкретные механизмы отбора абстрагируются и учитываются через обобщённые коэффициенты C_{read} и C_{merge} .

Этап вычисления релевантности документов предназначен для количественной оценки степени соответствия каждого документа из кандидатного множества пользовательскому запросу. В рамках модели релевантность определяется как аддитивная функция, отражающая соответствие содержания документа семантике пользовательского запроса [1, 3].

Релевантность документа d запросу q определяется следующим образом:

$$Rel(d, q) = \sum_{i=1}^{L_q} f(t_i, d) \quad (5)$$

где:

- L_q — количество терминов в пользовательском запросе.

- t_i — i -й термин запроса.
- $f(t_i, d)$ — функция вклада термина t_i в релевантность документа d .

Функция $f(t_i, d)$ может реализовываться с использованием статистических (например, TF-IDF), векторных (косинусное сходство) моделей [1], а также нейросетевых (эмбединги, трансформеры) и гибридных моделей [8]. Аддитивный характер функции означает независимый вклад каждого термина запроса в итоговую оценку документа. С точки зрения временной сложности, вычисление релевантности требует обработки каждого термина запроса для каждого документа из кандидатного множества, что приводит к зависимости:

$$T_{rel} = C_{rel} \times N_c \times L_q \quad (6)$$

где:

- C_{rel} — среднее время вычисления вклада одного термина в релевантность одного документа.
- N_c — количество документов в кандидатном множестве.
- L_q — количество терминов в пользовательском запросе.

Линейная зависимость обусловлена необходимостью вычисления функции $f(t_i, d)$ для каждой пары «термин — документ», при этом усложнение моделей релевантности увеличивает значение коэффициента C_{rel} без изменения вида зависимости. Формула релевантности не детализируется, поскольку модель ориентирована на оценку временных затрат.

Этап ранжирования предназначен для упорядочивания документов по убыванию релевантности и формирования итоговой поисковой выдачи, завершая вычислительную часть обработки запроса.

$$T_{rank} = C_{rank} \times N_c \times \log_2 K \quad (7)$$

Где:

- C_{rank} — среднее время операции сравнения.
- N_c — количество документов в кандидатном множестве.
- K — количество документов в итоговой выдаче.

Используется частичное ранжирование (top-K), при котором обрабатываются только наиболее релевантные документы, а логарифмический множитель отражает сложность операций поддержания структуры данных при их выборе [12].

Заключительный этап предназначен для формирования и передачи результата пользователю, включая подготовку метаданных и представление результатов поиска.

$$T_{resp} = C_{resp} \times K \quad (8)$$

где:

- C_{resp} — среднее время подготовки одного элемента выдачи.
- K — количество документов в списке.

Время этапа пропорционально размеру выдачи. Использование кеширования и оптимизированных форматов передачи данных снижает значение коэффициента C_{resp} и повышает воспринимаемую пользователем скорость работы системы.

Сравнительный анализ поисковых подходов

Ключевой поиск (Keyword-based search)

Ключевой поиск основан на совпадении терминов запроса с терминами документа без учета семантических связей [1]. Семантическая интерпретация запроса отсутствует:

$$C_{sem} = 0 \quad (9)$$

где C_{sem} — коэффициент вычислительных затрат на семантическую обработку.

Документы коллекции предварительно обрабатываются и индексируются с использованием инвертированного индекса, формируемого на офлайн-этапе и не учитываемого во времени отклика [1]. В процессе обработки запроса выполняются токенизация, нормализация, удаление стоп-слов и, при необходимости, лемматизация и сопоставление терминов запроса со словарем индекса.

Множество кандидатных документов формируется на основе постинговых списков терминов запроса, для каждого термина t_i извлекается список документов размера F_i . Итоговое множество кандидатов размера N_c получается посредством операций объединения или пересечения.

Оценка релевантности документов выполняется с использованием частотных статистических моделей, таких как TF-IDF [1] или BM25 [3]. В рамках теоретического расчёта время вычисления релевантности описывается выражением:

$$T_{rel}^{kw} = C_{rel}^{kw} \times N_c \times L_q \quad (10)$$

где:

- T_{rel}^{kw} — время вычисления релевантности документов в рамках ключевого поиска.
- C_{rel}^{kw} — коэффициент, отражающий вычислительную стоимость обработки одного документа

по одному термину запроса, включая подсчёт частот, вычисление весов и агрегацию частотных характеристик в итоговую оценку релевантности.

- N_c — количество документов в кандидатном множестве, для которых необходимо вычислить оценку релевантности.
- L_q — длина пользовательского запроса, выраженная числом терминов после лексической предобработки.

Таким образом, сложность этапа вычисления релевантности линейно зависит как от размера кандидатного множества, так и от длины запроса. Ключевой поиск отличается высокой вычислительной эффективностью и хорошей масштабируемостью, основным ограничением подхода является неспособность учитывать смысловые соответствия между запросом и документом при отсутствии прямого совпадения терминов.

Семантический поиск (Semantic search)

Семантический поиск основан на использовании распределённых векторных представлений текста [6]. В рамках данного подхода каждый документ преобразуется в вектор фиксированной размерности в латентном пространстве признаков с помощью нейросетевых моделей, например, Sentence-BERT [7]. Такие векторы отражают семантическое содержание текста и обеспечивают измерение смысловой близости между текстами независимо от конкретных словоформ.

Словарная интерпретация запроса отсутствует или несущественна, что в рамках теоретической модели выражается следующим образом:

$$C_{dict} \approx 0 \quad (11)$$

где C_{dict} — коэффициент вычислительных затрат, связанных с лексической интерпретацией запроса и обращением к словарным структурам.

Формирование кандидатного множества осуществляется путём поиска ближайших векторов документов в латентном пространстве. Для обеспечения приемлемого времени отклика используются структуры приближённого поиска ближайших соседей [9, 10, 11]. В рамках модели величины F_i интерпретируются как размеры локальных множеств ближайших соседей, возвращаемых алгоритмом поиска, а N_c — как итоговое число документов, отобранных по критерию семантической близости к запросу.

Вычисление релевантности документов основано на вычислении меры сходства между вектором запроса и векторами документов, как правило, с использованием косинусного сходства [1]. Временные затраты на данный этап описываются формулой:

$$T_{rel}^{sem} = C_{rel}^{sem} \times N_c \quad (12)$$

где:

- T_{rel}^{sem} — время вычисления релевантности документов в семантическом поиске.
- C_{rel}^{sem} — коэффициент, отражающий совокупную стоимость инференса нейросетевой модели и выполнения векторных операций (скалярных произведений или вычисления косинусной меры сходства).
- N_c — количество документов-кандидатов, для которых вычисляется мера семантического сходства.

В отличие от ключевого поиска, вычислительная сложность данного этапа практически не зависит от длины запроса L_q , поскольку запрос представлен одним вектором фиксированной размерности. Семантический поиск обеспечивает более высокое качество поиска и устойчивость к переформулировкам запроса, однако требует значительных вычислительных ресурсов и более сложной инфраструктуры [4, 5, 8].

Гибридный поиск (Hybrid search)

Гибридный поиск сочетает преимущества ключевого и семантического подходов и рассматривается как компромиссное решение, обеспечивающее баланс между качеством поиска и вычислительной эффективностью. Основная идея гибридного подхода заключается в использовании ключевого поиска для быстрого формирования ограниченного множества кандидатов с последующим семантическим уточнением результатов [8].

На этапе интерпретации запроса выполняется как традиционная лексическая обработка, так и построение семантического векторного представления, что в рамках теоретической модели отражается ненулевыми значениями обоих коэффициентов:

$$C_{dict} > 0, C_{sem} > 0 \quad (13)$$

где:

- C_{dict} — коэффициент вычислительных затрат лексической интерпретации запроса.
- C_{sem} — коэффициент вычислительных затрат семантической интерпретации запроса.

Формирование кандидатного множества осуществляется с использованием инвертированного индекса, аналогично ключевому поиску. В результате формируется кандидатное множество документов размером N_c , существенно меньшее по сравнению с полной коллекцией. Таким образом, применение вычислительно затратных семантических методов становится оправданным.

Вычисление релевантности в гибридном подходе включает два последовательных этапа и описывается выражением:

$$T_{rel}^{hyb} = C_{rel}^{kw} \times N_c \times L_q + C_{rel}^{sem} \times K, K \ll N_c \quad (14)$$

где:

- T_{rel}^{hyb} — суммарное время вычисления релевантности в гибридном поиске.
- C_{rel}^{kw} — коэффициент вычислительной стоимости ключевого этапа оценки релевантности.
- N_c — количество документов в кандидатном множестве, сформированном на этапе ключевого поиска.
- L_q — длина пользовательского запроса в терминах после лексической предобработки.
- C_{rel}^{sem} — коэффициент вычислительной стоимости семантического этапа оценки релевантности.
- K — количество документов верхней части ранжированного списка, для которых выполняется семантическое переранжирование.

Первое слагаемое ($C_{rel}^{kw} \times N_c \times L_q$) соответствует вычислению ключевой оценки релевантности для всех документов кандидатного множества. Данный этап полностью аналогичен этапу вычисления релевантности в ключевом поиске и обеспечивает предварительное ранжирование документов на основе совпадения терминов запроса. Второе слагаемое ($C_{rel}^{sem} \times K$) описывает этап семантического переранжирования, который применяется только к верхней части списка документов, содержащей K наиболее перспективных кандидатов. Здесь K выбирается существенно меньшим, чем N_c , что позволяет ограничить вычислительные затраты на инференс нейросетевой модели и векторные операции [8].

Данный подход позволяет сохранить высокую скорость ключевого поиска при одновременном повы-

шении качества ранжирования за счёт точечного применения семантической модели. Гибридный поиск обеспечивает высокое качество выдачи при контролируемой вычислительной сложности и является наиболее практически применимым подходом для систем поиска научных публикаций [4, 5].

Структурная схема обработки запроса в модели времени отклика

Структурная схема обработки запроса в модели времени отклика представлена на рис. 1. Схема отражает последовательность этапов обработки запроса в рамках аддитивной модели. Этапы предварительной обработки, интерпретации запроса и формирования множества кандидатных документов являются общими для всех подходов. Различия локализованы на этапе вычисления релевантности документов. Этапы ранжирования документов и формирования ответа унифицированы.

Параметризация коэффициентов модели времени отклика

Коэффициенты C интерпретируются как средние временные затраты, что позволяет учитывать вариативность исполнения операций, обусловленную особенностями аппаратной платформы и программной реализации. Компоненты времени рассматриваются в рамках аддитивной аппроксимации.

Коэффициент предварительной обработки запроса (C_{prep})

Коэффициент C_{prep} характеризует среднее время выполнения операций предварительной обработки одного термина пользовательского запроса. В состав данных операций входят: лексический анализ входной строки, нормализация словоформ и лемматизация, обеспечивающие приведение запроса к унифицированному пред-

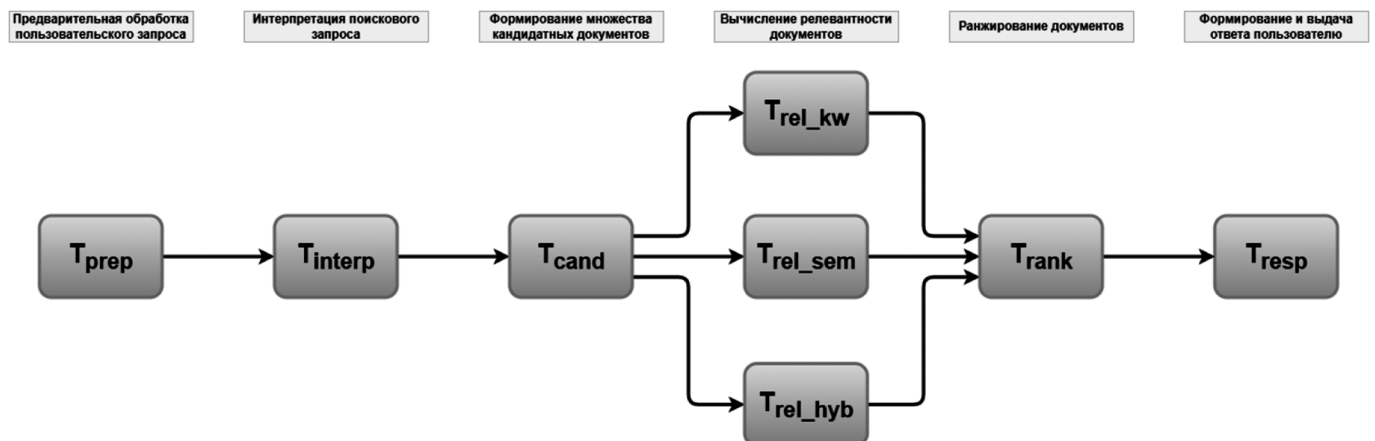


Рис. 1. Структурная схема обработки запроса в модели времени отклика для ключевого, семантического и гибридного поисковых подходов

ставлению, пригодному для последующей интерпретации.

$$C_{prep} = t_{CPU}^{prep} + t_{MEM}^{access} + t_{SW}^{over} \quad (15)$$

где:

- t_{CPU}^{prep} — среднее время выполнения процессорных инструкций, связанных с анализом и преобразованием одного термина. Значение определяется тактовой частотой процессора, числом доступных ядер и применяемыми оптимизациями (SIMD, векторизация).
- t_{MEM}^{access} — среднее время доступа к данным, размещённым в кэш-памяти и оперативной памяти (морфологические словари, таблицы соответствий, правила обработки).
- t_{SW}^{over} — накладные расходы программной платформы, обусловленные особенностями реализации (интерпретируемый код, вызовы функций библиотек, управление памятью).

Значение коэффициента C_{prep} уменьшается при использовании многопоточности и аппаратной векторизации, увеличивается для морфологически сложных языков вследствие возрастания числа операций анализа и обращений к словарным структурам.

Коэффициенты интерпретации запроса

Коэффициент поиска термина в словаре (C_{dict})

Коэффициент C_{dict} описывает среднее время выполнения операции поиска одного термина запроса в словарных и индексных структурах системы и отражает эффективность организации данных и используемых структур поиска.

$$C_{dict} = t_{CPU}^{search} + t_{MEM/DSK}^{access} + t_{SW}^{over} \quad (16)$$

где:

- t_{CPU}^{search} — среднее время выполнения операций сравнения и навигации по структурам данных (хэш-таблицы, деревья, словари).
- $t_{MEM/DSK}^{access}$ — среднее время доступа к данным словаря, размещённым в оперативной памяти или на дисковом носителе.
- t_{SW}^{over} — накладные расходы программного обеспечения, связанные с вызовами функций поиска и обработкой результатов.

Коэффициент семантической интерпретации термина (C_{sem})

Коэффициент C_{sem} отражает среднее время вычисления семантического векторного представления одного

термина запроса, включая получение эмбедингов и выполнение инференса семантических моделей.

$$C_{sem} = t_{CPU}^{sem} + t_{MEM}^{access} + t_{ACC}^{infer} \quad (17)$$

где:

- t_{CPU}^{sem} — среднее время выполнения процессорных вычислений, связанных с подготовкой данных.
- t_{MEM}^{access} — среднее время доступа к параметрам модели, хранящимся в оперативной памяти.
- t_{ACC}^{infer} — время инференса семантической модели на аппаратном ускорителе (ACC — аппаратный вычислительный ускоритель).

Коэффициент C_{sem} зависит от архитектуры используемой модели (Word2Vec, BERT, Sentence Transformers) и наличия специализированных вычислительных ускорителей.

Коэффициенты формирования множества кандидатов документов

Коэффициент чтения документов из индекса (C_{read})

Коэффициент C_{read} характеризует среднее время извлечения информации об одном документе из индексных структур системы.

$$C_{read} = t_{CPU}^{read} + t_{MEM/DSK}^{access} + t_{SW}^{over} \quad (18)$$

где:

- t_{CPU}^{read} — среднее время выполнения процессорных инструкций для извлечения элементов индекса.
- $t_{MEM/DSK}^{access}$ — задержка чтения данных из оперативной памяти или с диска.
- t_{SW}^{over} — накладные расходы программного обеспечения.

Коэффициент объединения индексных списков (C_{merge})

Коэффициент C_{merge} описывает среднее время объединения и фильтрации индексных списков, соответствующих терминам запроса.

$$C_{merge} = t_{CPU}^{merge} + t_{MEM}^{access} + t_{SW}^{over} \quad (19)$$

где:

- t_{CPU}^{merge} — среднее время выполнения операций объединения и фильтрации индексных списков.
- t_{MEM}^{access} — доступ к индексным структурам в оперативной памяти.
- t_{SW}^{over} — накладные расходы библиотек обработки данных.

Коэффициенты вычисления релевантности**Коэффициент частотной оценки релевантности (C_{rel}^{kw})**

Коэффициент C_{rel}^{kw} отражает средние временные затраты на вычисление оценок релевантности в частотных моделях (TF-IDF, BM25) для одного термина запроса и одного документа (временные затраты на одну операцию оценки релевантности пары «термин–документ»).

$$C_{rel}^{kw} = t_{CPU}^{rel,kw} + t_{MEM}^{access,kw} + t_{SW}^{over,kw} \quad (20)$$

где:

- $t_{CPU}^{rel,kw}$ — время вычисления частот, формирования весов и агрегации оценок.
- $t_{MEM}^{access,kw}$ — доступ к индексным и частотным структурам.
- $t_{SW}^{over,kw}$ — накладные расходы реализации.

Коэффициент семантической оценки релевантности (C_{rel}^{sem})

Коэффициент C_{rel}^{sem} описывает среднее время вычисления семантической близости между векторным представлением запроса и отдельным документом.

$$C_{rel}^{sem} = t_{CPU}^{rel,sem} + t_{MEM}^{access,sem} + t_{ACC}^{infer} + t_{SW}^{over,sem} \quad (21)$$

где:

- $t_{CPU}^{rel,sem}$ — операции подготовки данных и постобработки.
- $t_{MEM}^{access,sem}$ — доступ к эмбедингам документов.
- t_{ACC}^{infer} — инференс нейросетевой модели на вычислительном ускорителе.
- $t_{SW}^{over,sem}$ — накладные расходы программной платформы.

Коэффициент ранжирования документов (C_{rank})

Коэффициент C_{rank} характеризует среднее время выполнения одной операции сравнения документов в процессе сортировки по убыванию релевантности.

$$C_{rank} = t_{CPU}^{cmp} + t_{MEM}^{access} + t_{SW}^{sort} \quad (22)$$

где:

- t_{CPU}^{cmp} — среднее время одной операции сравнения.
- t_{MEM}^{access} — доступ к значениям релевантности документов.
- t_{SW}^{sort} — накладные расходы алгоритма сортировки.

Коэффициент формирования ответа пользователю (C_{resp})

Коэффициент C_{resp} отражает средние временные затраты на формирование и передачу информации об одном документе пользователю.

$$C_{resp} = t_{CPU}^{format} + t_{MEM}^{access} + t_{SW}^{serialize} + t_{NET}^{send} \quad (23)$$

где:

- t_{CPU}^{format} — среднее время формирования структуры ответа.
- t_{MEM}^{access} — доступ к данным документов.
- $t_{SW}^{serialize}$ — сериализация данных (JSON, XML).
- t_{NET}^{send} — передача ответа.

Компонент t_{NET}^{send} отражает внутренние затраты серверной части системы на подготовку и передачу результатов поиска между её функциональными модулями. Задержки, связанные с передачей данных по внешним сетям, в данный компонент не включаются.

Представленные коэффициенты отражают влияние аппаратных ресурсов и программной платформы на временные затраты ключевых операций обработки поискового запроса [1, 13].

Экспериментальное исследование и оценка полученных результатов**Методика проведения эксперимента**

Экспериментальное исследование проводилось в локальной вычислительной среде без использования сетевых соединений, что позволило исключить влияние сетевых задержек и анализировать алгоритмическую составляющую времени отклика. Аппаратная конфигурация: процессор Intel Core i5, 8 Гб оперативной памяти. Программная среда: Python 3.10, NumPy.

Использовался синтетический корпус документов, имитирующий структуру базы научных публикаций: 6000 документов, словарь — 5000 терминов, средняя длина документа — 80 терминов, размерность векторных представлений — 64, длина запроса — 8 терминов, глубина выдачи — 10 документов. В гибридной модели выполнялся предварительный отбор 50 документов для семантического переранжирования. Равномерное распределение терминов исключило влияние содержательной специфики.

Для каждого подхода использовался фиксированный запрос длиной 8 терминов. Выполнялось 30 измерений после исключения первых 5 запусков (прогрев среды), измерение производилось посредством инструментального профилирования программной реализации. Теоретическая оценка вычислялась для каждого запуска на основе измеренных компонентов временных затрат, что обеспечило проверку аддитивной структуры модели.

Результаты эксперимента

Произведено сопоставление теоретических и экспериментальных значений времени выполнения запроса. Сводные данные представлены в табл. 1.

Таблица 1.

Сравнение теоретического и экспериментального времени отклика

Поисковый подход	$T_{\text{теор}}$ (мс)	$T_{\text{эксп}}$ (мс)	Абсолютная разность (мс)	Погреш- ность
Ключевой поиск	0,358	0,360	0,002	0,65 %
Семантический поиск	3,740	3,742	0,002	0,06 %
Гибридный поиск	0,508	0,628	0,120	19,09 %

Для ключевого и семантического поиска теоретические и экспериментальные значения согласуются. Для гибридного подхода относительная погрешность достигает 19 %, что обусловлено дополнительными накладными расходами интерпретируемой среды, повторным формированием структур данных и промежуточными операциями.

Заключение

В работе выполнена формализация процесса обработки поискового запроса и разработана аналитическая модель времени отклика для ключевого, семантическо-

го и гибридного подходов, адекватно описывающая зависимость временных затрат от структуры алгоритма. Для ключевого и семантического поиска расхождение между теоретическими и экспериментальными значениями не превышает 1 % (0,65 % и 0,06 % соответственно). Для гибридного поиска наблюдается повышенная относительная погрешность ($\approx 19\%$), обусловленная дополнительными накладными расходами интерпретируемой среды, при малом абсолютном отклонении (0,120 мс).

Установлено, что ключевой поиск обеспечивает минимальное время формирования выдачи ($\approx 0,36$ мс), семантический является наиболее ресурсоёмким ($\approx 3,7$ мс), а гибридный ($\approx 0,63$ мс) занимает промежуточное положение. Семантическая обработка выполняется примерно в 10 раз медленнее ключевой и в 6 раз медленнее гибридной, тогда как гибридный поиск приблизительно в 1,7 раза медленнее ключевого поиска, но более чем в 5 раз быстрее полной семантической обработки.

Полученные результаты подтверждают применимость модели для оценки времени отклика и обоснования выбора поискового подхода с учётом требований к производительности.

ЛИТЕРАТУРА

1. Маннинг К., Рагхаван П., Шютце Х. Введение в информационный поиск пер. с англ. — М.: Вильямс, 2011. — 528 с.
2. Nielsen J. Usability Engineering. — San Francisco, CA: Morgan Kaufmann, 1994. — 362 p.
3. Robertson S., Zaragoza H. The Probabilistic Relevance Framework: BM25 and Beyond // Foundations and Trends in Information Retrieval. — 2009. — Vol. 3, No. 4. — p. 333–389.
4. Järvelin K., Kekäläinen J. Cumulated Gain-Based Evaluation of IR Techniques // ACM Transactions on Information Systems. — 2002. — Vol. 20, No. 4. — p. 422–446.
5. Moffat A., Zobel J. Rank-Biased Precision for Measurement of Retrieval Effectiveness // ACM Transactions on Information Systems. — 2008. — Vol. 27, No. 1. — Article 2.
6. Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space. — arXiv:1301.3781. — 2013.
7. Reimers N., Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks // Proceedings of EMNLP-IJCNLP 2019. — 2019. — p. 3982–3992.
8. Mitra B., Craswell N. An Introduction to Neural Information Retrieval // Foundations and Trends in Information Retrieval. — 2018. — Vol. 13, No. 1. — p. 1–126.
9. Muja M., Lowe D.G. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration // Proceedings of VISAPP 2009. — 2009. — p. 331–340.
10. Malkov Y.A., Yashunin D.A. Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2020. — Vol. 42, No. 4. — p. 824–836.
11. Johnson J., Douze M., Jégou H. Billion-scale Similarity Search with GPUs // IEEE Transactions on Big Data. — 2021. — Vol. 7, No. 3. — p. 535–547.
12. Кормен Т.Х., Лейзерсон Ч.Э., Ривест Р.Л., Штайн К. Алгоритмы: построение и анализ. — 3-е изд. — М.: Вильямс, 2013. — 1328 с.
13. Таненбаум Э., Остин Т. Архитектура компьютера. — 6-е изд. — СПб.: Питер, 2013. — 816 с.

© Горячкин Борис Сергеевич (bsgor@mail.ru); Сыса Павел Валерьевич (sysa.pavel@yandex.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

ОБЗОР МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ПОИСКА АНОМАЛИЙ В РАСПРЕДЕЛЕННЫХ СИСТЕМАХ

REVIEW OF ARTIFICIAL INTELLIGENCE METHODS FOR DETECTING ANOMALIES IN DISTRIBUTED COMPUTERS

**D. Gritsenko
K. Parfentiev**

Summary. This article presents an overview of modern artificial intelligence (AI) methods used to detect anomalies and failures in distributed information systems (IS). It examines the scale of the problem of monitoring large software systems and the evolution of incident management approaches — from traditional threshold rules to the AIOps (Artificial Intelligence for IT Operations) paradigm based on machine learning (ML). The application of deep learning, natural language processing (NLP), and time series methods to the analysis of unstructured event logs is analyzed. Special attention is paid to the formalization of the anomaly detection problem, the role of monitoring in continuous integration/continuous deployment (CI/CD) processes, and criteria for evaluating the effectiveness of algorithms. A conclusion is drawn regarding the potential of hybrid approaches combining statistical and neural network models.

Keywords: AIOps, anomaly detection, distributed systems, machine learning, log analysis, monitoring, neural networks, incident management, CI/CD, deep learning.

Гриценко Дмитрий Андреевич

аспирант, Московский государственный
университет им. Н.Э. Баумана
dimgrit@mail.ru

Парфентьев Кирилл Викторович

Кандидат технических наук, доцент, Московский
государственный университет им. Н.Э. Баумана
parfentiev@bmstu.ru

Аннотация. В статье представлен обзор современных методов искусственного интеллекта (ИИ), применяемых для обнаружения аномалий и сбоев в распределенных информационных системах (ИС). Рассматриваются масштаб проблемы мониторинга крупных программных комплексов и эволюция подходов к управлению инцидентами — от традиционных пороговых правил до парадигмы AIOps (Artificial Intelligence for IT Operations — AIOps), основанной на машинном обучении (МО). Анализируется применение методов глубокого обучения, обработки естественного языка (Natural Language Processing — NLP) и временных рядов для анализа неструктурированных журналов событий (логов). Отдельное внимание уделено формализации задачи поиска аномалий, роли мониторинга в процессах непрерывной интеграции и развертывания (Continuous Integration / Continuous Deployment — CI/CD), а также критериям оценки эффективности алгоритмов. Сделан вывод о перспективности гибридных подходов, сочетающих статистические и нейросетевые модели.

Ключевые слова: AIOps, обнаружение аномалий, распределенные системы, машинное обучение, анализ логов, мониторинг, нейронные сети, управление инцидентами, CI/CD, глубокое обучение.

Введение

Современные информационные системы (ИС) переживают период радикальной трансформации архитектурных решений. Монолитные приложения уступают место распределённым микросервисным архитектурам, развёрнутым в гибридных облачных средах. Согласно отчёту Cloud Native Computing Foundation (CNCF), к 2024 году более 96 % крупных организаций используют контейнерные технологии в продуктивных средах, а среднее число микросервисов в составе одной системы превышает несколько сотен [1]. Подобная архитектурная сложность приводит к экспоненциальному росту объёма телеметрических данных — метрик, журналов событий и трассировок (logs, metrics, traces), что делает задачу мониторинга качественно новой проблемой.

Масштаб проблемы определяется не только объёмом данных, но и стоимостью простоев. По данным аналитического агентства Gartner, средняя стоимость

одной минуты простоя критически важной ИС в крупном предприятии составляет от 5,6 до 9 тыс. долларов США [2]. При этом, согласно исследованию IBM Global Services, до 70 % инцидентов в распределённых системах связаны не с отказом аппаратного обеспечения, а с программными аномалиями, ошибками конфигурации и каскадными сбоями, выявление которых при помощи традиционных средств мониторинга затруднено [3]. Дополнительную сложность создаёт скорость выпуска изменений: в практике непрерывного развёртывания (Continuous Deployment) частота выкатываний обновлений в крупных интернет-компаниях достигает тысяч в сутки, что требует автоматизированных средств обнаружения деградаций.

В этих условиях формируется самостоятельная область знаний — искусственный интеллект для эксплуатации ИТ (Artificial Intelligence for IT Operations — AIOps). Термин, введённый Gartner в 2017 году, обозначает комплекс методов, использующих машинное обучение (МО),

статистический анализ и инженерии больших данных для автоматизации операционных процессов: выявления аномалий, корреляции событий, прогнозирования отказов и поиска первопричин (Root Cause Analysis — RCA) [4]. По оценкам аналитиков, мировой рынок AIOps-решений демонстрирует среднегодовой темп роста (CAGR) на уровне 25 % и должен достичь 32 млрд долларов США к 2028 году.

Целью настоящей работы является систематизация существующих методов искусственного интеллекта (ИИ), применяемых для обнаружения аномалий и сбоев в распределённых ИС, с акцентом на анализ неструктурированных журналов событий. Для достижения цели последовательно решаются задачи: проводится обзор традиционных и интеллектуальных методов мониторинга; формализуется задача поиска аномалий в логах как задача обработки последовательностей; обсуждаются критерии оценки качества алгоритмов; выполняется сравнительный анализ известных решений.

1. Обзор существующих методов обнаружения сбоев и управления инцидентами

1.1. Традиционные статистические и пороговые методы мониторинга ИС

Исторически первый класс методов мониторинга основан на сравнении наблюдаемых значений метрик с предопределёнными пороговыми значениями. Подобный подход реализуется в большинстве классических систем — Nagios (USA), Zabbix LLC (Latvia), Prometheus, Grafana Labs (USA, New York). Несмотря на простоту, фиксированные пороги обладают рядом фундаментальных ограничений: они не учитывают сезонные колебания нагрузки, требуют ручной настройки сотен правил, генерируют большое число ложных срабатываний (false positives) и не способны обнаруживать ранее неизвестные типы аномалий [5].

Развитием порогового подхода являются адаптивные статистические методы, формализующие задачу обнаружения аномалий как проверку гипотезы о принадлежности нового наблюдения к статистическому распределению нормальной работы. Среди широко применяемых моделей следует выделить экспоненциальное сглаживание Holt–Winters, авторегрессионные интегрированные модели скользящего среднего (Autoregressive Integrated Moving Average — ARIMA), сезонные SARIMA-модели, а также методы на основе z-оценок и межквартильного размаха. Для одномерного временного ряда x_t условие аномалии может быть формализовано следующим образом: наблюдение признаётся аномальным, если $|x_t - \mu_t| > k \cdot \sigma_t$, где μ_t и σ_t — локальные оценки среднего и стандартного отклонения, а k — пороговый коэффициент чувствительности.

Преимуществом статистических методов является их интерпретируемость и низкая вычислительная сложность, что делает их пригодными для потоковой обработки на ребре сети. Однако они опираются на предположения о стационарности и гауссовости распределений, которые нарушаются в условиях многомерных коррелированных метрик распределённых систем. По данным сравнительных исследований [5, 6], точность одномерных статистических детекторов в реальных производственных средах редко превышает 60–70 % по метрике F1, что недостаточно для автоматизированной эксплуатации.

1.2. Применение методов машинного обучения и искусственного интеллекта (AIOps) для анализа логов

Парадигма AIOps предполагает использование методов МО и глубокого обучения (Deep Learning — DL) для автоматизации трёх ключевых задач: обнаружения аномалий, корреляции событий и поиска первопричин. В рамках задачи анализа журналов событий выделяют несколько технологических подходов, эволюционировавших в последнее десятилетие.

Первый класс методов основан на классических алгоритмах обучения без учителя. К ним относятся методы кластеризации (k-means, DBSCAN), модели одноклассового обучения (One-Class SVM), изолирующие леса (Isolation Forest) и автоэнкодеры. Их применение требует предварительного этапа извлечения структуры логов — парсинга, обеспечивающего преобразование неструктурированных текстовых сообщений в шаблоны (templates). Широкое распространение получили парсеры Drain, Spell и LogPAI, обеспечивающие точность извлечения шаблонов на уровне 0,9–0,97 на эталонных наборах данных LogHub [7].

Второй класс методов использует нейронные сети для моделирования последовательностей событий. Знаковой работой стала модель DeepLog (Du и др., 2017), представляющая последовательность шаблонов логов как поток токенов и обучающая рекуррентную нейронную сеть с долгой краткосрочной памятью (Long Short-Term Memory — LSTM) предсказывать следующий шаблон. Отклонение фактически наблюдаемого шаблона от прогноза, превышающее заданный порог top-g , классифицируется как аномалия [8]. Развитием идеи стали модели LogAnomaly (использующая семантические векторные представления `template2vec`), LogRobust (применяющая внимание на основе Bi-LSTM) и LogBERT (Guo и др., 2021), реализующая принцип маскированного языкового моделирования из BERT для логов [9].

Третий класс представлен трансформерными архитектурами и большими языковыми моделями (Large

Language Models — LLM). Современные исследования демонстрируют, что предобученные на естественном языке модели семейства GPT, BERT и T5 после дообучения на доменных корпусах журналов показывают качество обнаружения аномалий, превосходящее специализированные нейросетевые архитектуры предыдущего поколения. В частности, в работе [10] сообщается о значениях F1, достигающих 0,98–0,99 на наборах данных HDFS и BGL при использовании трансформерных энкодеров.

В таблице 1 приведено сравнение представительных методов обнаружения аномалий по ряду качественных характеристик.

Таблица 1.

Сравнение методов обнаружения аномалий в логах

Метод / модель	Класс алгоритмов	Входные данные	F1 на HDFS
1	2	3	4
Isolation Forest	Обучение без учителя	Векторы признаков	0,78
DeepLog (LSTM)	Глубокое обучение	Последовательности шаблонов	0,96
LogAnomaly	Глубокое обучение	Семантические векторы	0,96
LogBERT	Трансформеры	Маскированные токены	0,98
GPT-fine-tuned	Большие языковые модели	Текст логов	0,99

1.3. Роль мониторинга в управлении релизными циклами и процессах непрерывной интеграции/развёртывания (CI/CD)

Внедрение практик непрерывной интеграции и непрерывного развёртывания (Continuous Integration / Continuous Deployment — CI/CD) кардинально изменило требования к системам мониторинга. Высокая частота развёртываний, многократно превышающая возможности ручного контроля качества, требует автоматизированных «контрактов наблюдаемости» (Service Level Objectives — SLO), в рамках которых система обнаружения аномалий выступает обязательным контролируемым гейтом в конвейере поставки.

В современной практике DevOps-команд выделяют три точки интеграции методов ИИ-мониторинга в CI/CD-конвейер. Во-первых, на этапе пререлизного тестирования применяются модели прогнозирования регрессий производительности, сравнивающие распределение метрик нагрузочных тестов нового релиза с эталонным. Во-вторых, в момент развёртывания используются

стратегии канареечных и сине-зелёных развёртываний с автоматическим анализом аномалий: если в течение наблюдательного окна интеллектуальный детектор фиксирует статистически значимое отклонение метрик новой версии от текущей, выполняется автоматический откат. В-третьих, на этапе эксплуатации мониторинг обеспечивает обратную связь, питающую процессы инцидент-менеджмента и постмортемного анализа.

Канонической реализацией данной концепции является подход progressive delivery, поддержанный платформами Argo Rollouts, Flagger и Spinnaker. Эмпирические данные [4, 6] показывают, что интеграция методов МО в процесс канареечного развёртывания снижает среднее время восстановления (Mean Time To Recovery — MTTR) на 30–55 % по сравнению с ручным контролем, а число неуспешных откатов уменьшается в 2–3 раза.

2. Формализация задачи поиска аномалий и критерии оценки эффективности

2.1. Формализация задачи поиска аномалий в неструктурированных логах как задачи обработки последовательностей

Журнал событий распределённой ИС представляет собой неструктурированный поток текстовых сообщений, генерируемых множеством слабосвязанных микросервисов. Формально журнал можно описать как последовательность $L = \{l_1, l_2, \dots, l_n\}$, где каждое сообщение $l_i = (t_i, s_i, c_i)$ состоит из временной метки t_i , идентификатора источника s_i и текстового содержания c_i . Прямое использование сырых текстов затруднено из-за высокой вариативности численных параметров и идентификаторов, поэтому первый этап обработки — парсинг логов — состоит в извлечении инвариантной структуры (шаблона) $e_i \in E$ и выделении переменных параметров.

После парсинга задача обнаружения аномалий формализуется как задача обработки последовательностей шаблонов. Пусть $S = (e_{(1)}, e_{(2)}, \dots, e_{(n)})$ — временное окно длины T последовательных событий. Требуется построить функцию $f: S \rightarrow \{0, 1\}$, относящую окно к классу «нормальное» (0) или «аномальное» (1). В рамках обучения без учителя функция строится через моделирование условного распределения $P(e_t | e_{(t-1)}, \dots, e_{(t-k)})$ с порогом лог-правдоподобия в качестве критерия аномалии. В рамках обучения с учителем оптимизируется функция потерь, например, бинарная кросс-энтропия:

$$L(\theta) = -\sum [y_i \cdot \log p_{\theta(x_i)} + (1 - y_i) \cdot \log (1 - p_{\theta(x_i)})], \quad (1)$$

где θ — обучаемые параметры модели, y_i — истинная метка аномальности окна, $p_{\theta(x_i)}$ — оценённая моделью вероятность аномалии. Подобная постановка естественным образом сводится к задачам моделирования языка,

обработке временных рядов и графов вызовов, что и обуславливает применимость глубоких нейросетевых архитектур.

Существенной особенностью задачи является сильный дисбаланс классов: доля аномальных событий в реальных журналах редко превышает 1 %, что требует применения специальных техник — взвешивания классов, контрастного обучения, генерации синтетических аномалий и обучения по типу одноклассовой классификации. Дополнительные сложности связаны с дрейфом концепта (concept drift), вызванным эволюцией шаблонов логов при выпуске новых версий программного обеспечения.

2.2. Выбор критериев оценки эффективности разрабатываемых методов и алгоритмов

Оценка качества методов обнаружения аномалий в распределённых ИС требует комплексной системы метрик, отражающих как корректность классификации, так и эксплуатационные свойства алгоритма. В качестве базовых метрик применяются классические показатели бинарной классификации — точность (Precision), полнота (Recall) и F1-мера, рассчитываемые на основе значений истинно-положительных (TP), ложно-положительных (FP), истинно-отрицательных (TN) и ложно-отрицательных (FN) результатов:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}), \text{Recall} = \text{TP} / (\text{TP} + \text{FN}), \\ \text{F1} = 2 \cdot \text{Precision} \cdot \text{Recall} / (\text{Precision} + \text{Recall}). \quad (2)$$

Однако только их использования недостаточно. Для задач с сильным дисбалансом классов рекомендуется использовать площадь под кривой полнота–точность (Area Under Precision-Recall Curve — AUPRC) как более устойчивую метрику по сравнению с традиционной AUROC. В условиях потокового мониторинга существенную роль играют темпоральные метрики: задержка обнаружения (detection latency), измеряемая как промежуток времени между фактическим возникновением аномалии и её фиксацией, а также частота ложных срабатываний на единицу времени (False Alarm Rate — FAR), напрямую влияющая на «усталость от уведомлений» (alert fatigue) оперативных служб.

Эксплуатационные критерии включают: вычислительную сложность модели (latency инференса и пропускная способность), требования к памяти, способность к инкрементному обучению (online learning) и адаптации к дрейфу данных, интерпретируемость прогнозов (explainability). Последний критерий приобретает особое значение в свете задачи поиска первопричин: метод должен не только сигнализировать о наличии аномалии, но и указывать на конкретный микросервис или фрагмент журнала, ответственный за деградацию.

Современные подходы, основанные на методах SHAP, LIME и механизмах внимания (attention), частично решают эту задачу [9, 10].

Сводный критерий пригодности метода для промышленного применения, в соответствии с подходом [6], предлагается формировать как взвешенную свёртку нормированных значений ключевых метрик качества и производительности с весовыми коэффициентами, отражающими приоритеты конкретной эксплуатационной среды.

Заключение

В работе выполнен систематический обзор современных методов искусственного интеллекта, применяемых для обнаружения аномалий в распределённых информационных системах. Показано, что традиционные пороговые и одномерные статистические методы, исторически образующие основу систем мониторинга, не справляются с масштабом и сложностью современных микросервисных архитектур, что подтверждается ограниченным значением F1-меры в реальных средах и высоким уровнем ложных срабатываний.

Парадигма AIOps, опирающаяся на методы машинного и глубокого обучения, обеспечивает качественно более высокий уровень автоматизации эксплуатационных процессов. Эволюция нейросетевых архитектур — от рекуррентных моделей семейства DeepLog к моделям семейства BERT и большим языковым моделям — позволила приблизить значение F1 для эталонных наборов данных к 0,98–0,99. Одновременно с этим в практике CI/CD-конвейеров формируется новая роль интеллектуального мониторинга как обязательного контрольного гейта в процессе непрерывного развёртывания, обеспечивающего автоматический откат проблемных версий.

Формализация задачи поиска аномалий в неструктурированных логах как задачи обработки последовательностей шаблонов открывает прямой путь для применения современных языковых моделей. Вместе с тем, такие практические вызовы, как сильный дисбаланс классов, дрейф концепта и необходимость интерпретации прогнозов, требуют дальнейших исследований. Перспективным направлением представляется разработка гибридных подходов, сочетающих преимущества статистических методов (вычислительная экономность, интерпретируемость) и нейросетевых моделей (точность, способность выявлять сложные паттерны), а также интеграция методов поиска первопричин с механизмами объяснимого ИИ. Сформулированный в работе набор критериев оценки эффективности может служить основой для построения сравнительной методики и обоснованного выбора метода для конкретной эксплуатационной среды.

ЛИТЕРАТУРА

1. Cloud Native Computing Foundation. CNCF Annual Survey 2023. San Francisco: CNCF. 2024. 56 p. URL: cncf.io/reports/cncf-annual-survey-2023/.
2. Гартнер Дж. The cost of IT downtime. Stamford: Gartner Research. 2022. URL: gartner.com/en/documents/3980412.
3. Иванов А.В., Петров С.К. Анализ причин отказов в распределенных программных системах. Программная инженерия. 2021. Т. 12. № 4. С. 175–186.
4. Сидорова Е.Н., Кузнецов М.А. Обзор технологий AIOps для управления крупными информационными системами. Инженерный вестник Дона. 2022. № 7. URL: ivdon.ru/ru/magazine/archive/n7y2022/7821.
5. Соколов В.П., Морозов Д.А. Сравнительный анализ алгоритмов обнаружения аномалий во временных рядах метрик мониторинга. Программные продукты и системы. 2020. № 3. С. 412–421.
6. Васильев И.С., Никитин А.Б. Методы интеллектуального мониторинга в облачных средах. Москва: ДМК Пресс. 2021. 384 с.
7. He P., Zhu J., Zheng Z., Lyu M.R. Drain: An online log parsing approach with fixed depth tree. Proceedings of the IEEE International Conference on Web Services (ICWS). Honolulu: IEEE. 2017. Pp. 33–40.
8. Du M., Li F., Zheng G., Srikanth V. DeepLog: Anomaly detection and diagnosis from system logs through deep learning. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM. 2017. Pp. 1285–1298.
9. Guo H., Yuan S., Wu X. LogBERT: Log Anomaly Detection via BERT. Proceedings of the International Joint Conference on Neural Networks (IJCNN). Shenzhen: IEEE. 2021. Pp. 1–8.
10. Le V.H., Zhang H. Log-based anomaly detection with deep learning: how far are we? Proceedings of the 44th International Conference on Software Engineering (ICSE). Pittsburgh: ACM. 2022. Pp. 1356–1367.
11. Захаров А.А., Лебедев К.С. Применение нейронных сетей трансформерной архитектуры для анализа журналов событий. Вестник компьютерных и информационных технологий. 2023. № 5. С. 28–36.
12. Громов В.Г. Управление инцидентами в практиках DevOps и SRE: исследование подходов. Открытые системы. СУБД. 2022. № 2. С. 18–24.

© Гриценко Дмитрий Андреевич (dimgrit@mail.ru); Парфентьев Кирилл Викторович (parfentiev@bmstu.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

ИССЛЕДОВАНИЕ АРХИТЕКТУРЫ СИСТЕМЫ ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА АРКТИКИ

RESEARCH OF THE ARCHITECTURE ENVIRONMENTAL MONITORING SYSTEM OF THE ARCTIC

O. Evdoshenko

Summary. The article is devoted to the development of an architecture an intelligent system environmental monitoring to Arctic designed to collect, process and analyze data on anthropogenic impacts on the climatic and environmental parameters of the region in conditions of extreme climatic factors and limited infrastructural accessibility. A three-level hierarchical model is proposed, including autonomous sensor modules (Level 1), aggregating nodes with peripheral data processing functions (Level 2) and central gateways providing information transmission to the cloud infrastructure via satellite communication channels (Level 3). Special attention is paid to the adaptation of architecture to Arctic specifics: low temperatures, icing, geomagnetic disturbances and the need for long-term autonomous operation.

Keyword: Arctic, environmental monitoring, architecture of an intelligent system.

Евдошенко Олег Игоревич

кандидат технических наук, доцент, МИРЭА —
Российский технологический университет (г. Москва)
evdoshenko@mirea.ru

Аннотация. Статья посвящена разработке архитектуры интеллектуальной системы экологического мониторинга Арктики, предназначенной для сбора, обработки и анализа данных об антропогенном воздействии на климатические и экологические параметры региона в условиях экстремальных природно-климатических факторов и ограниченной инфраструктурной доступности. Предложена трехуровневая иерархическая модель, включающая автономные сенсорные модули (Уровень 1), агрегирующие узлы с функциями периферийной обработки данных (Уровень 2) и центральные шлюзы, обеспечивающие передачу информации в облачную инфраструктуру через спутниковые каналы связи (Уровень 3). Особое внимание уделено адаптации архитектуры к арктической специфике: низким температурам, обледенению, геомагнитным возмущениям и необходимости длительной автономной работы.

Ключевые слова: Арктика, экологический мониторинг, архитектура интеллектуальной системы.

Введение

Арктика является одной из самых уязвимых и быстро меняющихся климатических зон планеты. В последние десятилетия на фоне глобального потепления и растущей хозяйственной активности в регионе возрастает нагрузка на хрупкую арктическую экосистему [1]. Освоение месторождений полезных ископаемых, интенсификация судоходства, расширение инфраструктурных проектов, а также воздействие загрязняющих веществ приводят к ускоренной трансформации природной среды, угрожая биологическому разнообразию, традиционному укладу жизни коренных народов и глобальному климатическому балансу.

Одной из ключевых проблем современного этапа экологического управления в Арктике является недостаточная оперативность и точность сбора данных об антропогенных воздействиях, а также ограниченные возможности их интерпретации и анализа. Традиционные методы мониторинга не позволяют в полной мере охватывать территорию, характеризующуюся экстремальными погодными условиями, труднодоступностью и отсутствием устойчивой инфраструктуры. В этих условиях особое значение приобретает внедрение интеллектуальных автоматизированных систем, сочетающих

возможности Интернета вещей (IoT) и технологий искусственного интеллекта (ИИ).

Цель данной статьи: провести анализ архитектуры системы экологического мониторинга Арктики и описание уровней данной системы.

Общее представление об архитектуре системы

Архитектура системы комплексного экологического мониторинга Арктики базируется на трехуровневой иерархической модели, оптимизированной для работы в условиях экстремальных арктических условий при жестких ограничениях энергопотребления [2]. Данная модель обеспечивает сбор, первичную обработку и передачу данных об антропогенном воздействии на климатические и экологические параметры региона.

Описание уровней иерархической модели

Конечные устройства (Уровень 1) представляют собой массовые автономные сенсорные модули, размещаемые в наземных и надводных конфигурациях [3]. Их функция ограничивается сбором первичных метеорологических параметров (температура, влажность, атмосферное давление, скорость и направление ветра,

солнечная радиация), показателей состояния вечной мерзлоты (температура и влажность грунта на различных глубинах, деформации), характеристик снежно-ледового покрова (высота снега, температура поверхности, альбедо), базовых показателей качества воды (температура, мутность, электропроводность, растворенный кислород), а также концентраций ключевых газовых компонентов (CO_2 , CH_4) с использованием специализированных низкоэнергетических сенсоров. Для обеспечения длительной автономной работы (не менее 1 года) применяются стратегии глубокого сна микроконтроллеров, прерывистый режим активации энергоемких датчиков и передача данных через LPWAN-протоколы (LoRaWAN, NB-IoT) или подводную акустическую связь с интервалами 15–60 минут. Плотность размещения варьируется от 0.5–3 км в зонах интенсивного антропогенного воздействия (судоходные пути, промышленные объекты) до 5–20 км на фоновых территориях, что определяется пространственной изменчивостью контролируемых параметров и рекомендациями международных организаций (WMO, Arctic Council).

Агрегирующие узлы (Уровень 2) выполняют функции концентраторов данных от групп устройств Уровня 1 в радиусе до 10–15 км для наземных систем и до 2 км для подводных конфигураций. Эти узлы обладают повышенными вычислительными ресурсами, позволяющими осуществлять базовый контроль качества данных (выявление аномалий, проверка допустимых диапазонов), первичную агрегацию, сжатие информации и буферизацию при потере связи с верхним уровнем. Они также управляют режимами работы подчиненных L1-устройств (корректировка периодичности опроса, инициирование калибровочных процедур) и могут размещать более точные, но энергозатратные сенсоры (миниатюрные аэталометры для черного углерода, спектрометры для газов, системы очистки датчиков воды). Расстояние между узлами L2 составляет 15–25 км в приоритетных зонах и 30–50 км на остальных территориях, обеспечивая необходимое перекрытие зон ответственности для отказоустойчивости. Энергоснабжение реализуется через гибридные системы (солнечные панели, термогенераторы, аккумуляторы большой емкости).

Центральные шлюзы (Уровень 3) выступают точками консолидации информации от множества узлов L2 (в радиусе 70–120 км) и обеспечивают надежную передачу данных в облачную инфраструктуру через спутниковые каналы связи (Iridium Certus, в перспективе — Starlink). Обладая максимальными вычислительными мощностями в периферийной сети, они выполняют окончательное пакетирование данных, реализуют продвинутые алгоритмы маршрутизации, осуществляют ко-

ординацию работы всей сети (диагностика узлов, обновление ПО) и долгосрочное кэширование информации. Размещение L3-шлюзов привязано к пунктам логистической инфраструктуры (научные станции, аэропорты) для упрощения обслуживания. Обеспечение их энергоснабжения требует наиболее мощных и стабильных решений.

Интеграция данных со всех уровней, их углубленный анализ, хранение в БД, запуск прогностических моделей (включая алгоритмы машинного обучения) и визуализация осуществляются исключительно в облачной среде. Система проектируется с соблюдением принципов интероперабельности через использование открытых стандартов (OGC SensorThings API, MQTT, CoAP) и обеспечивает механизмы дистанционной калибровки, верификации данных и адаптивного управления режимами мониторинга в ответ на изменение условий окружающей среды. Учет специфики Арктики (низкие температуры, обледенение, ограниченная доступность, геомагнитные возмущения) заложен на всех этапах проектирования архитектуры.

Заключение

В условиях ускоренного изменения арктической среды и возрастающего антропогенного давления эффективный экологический мониторинг региона требует перехода от разрозненных наблюдений к интегрированным автоматизированным решениям. Предложенная в статье трехуровневая иерархическая архитектура системы экологического мониторинга, основанная на сочетании технологий Интернета вещей и искусственного интеллекта, позволяет преодолеть ключевые ограничения традиционных методов сбора данных, связанные с экстремальными климатическими условиями, труднодоступностью территорий и дефицитом энергоресурсов.

Распределенная структура, включающая автономные сенсорные модули, агрегирующие узлы и центральные шлюзы, обеспечивает гибкость масштабирования, отказоустойчивость и адаптивность системы. Использование LPWAN-протоколов, гибридных источников энергоснабжения и спутниковых каналов связи гарантирует устойчивую работу оборудования в условиях низких температур, обледенения и отсутствия развитой наземной инфраструктуры. Вынос вычислительных задач на периферийные узлы и последующая глубокая аналитика в облачной среде с применением алгоритмов машинного обучения позволяют не только оперативно выявлять аномалии и оценивать антропогенные воздействия, но и прогнозировать изменения климатических и экологических параметров.

ЛИТЕРАТУРА

1. Захаров В. М. Экологические угрозы и устойчивое развитие Арктического региона [Текст] — М.: Наука, 2021. — 224 с.
2. Бакланов А.А., Дроздов А.В. Моделирование климатических процессов в северных широтах [Текст] — СПб.: Гидрометеиздат, 2020. — 148 с.
3. Кудрявцев В.А. Сенсорные сети и устойчивость в полярных условиях [Текст] — М.: МГТУ им. Баумана, 2022. — 132 с.
4. Обработка сигналов от датчиков в полевых условиях / Гончаров А.Ю. — Красноярск: СФУ, 2021. — 101 с.
5. Работа IoT-сенсоров в условиях обледенения / Чернов В.М. // Полярная техника. — 2023. — № 3. — С. 53–59.

© Евдошенко Олег Игоревич (evdoshenko@mirea.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

МНОГОЛЕТНИЙ АНАЛИЗ И ПРЕДИКТИВНОЕ МОДЕЛИРОВАНИЕ КЛИМАТИЧЕСКИХ ИЗМЕНЕНИЙ В МОСКОВСКОМ РЕГИОНЕ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

LONG-TERM ANALYSIS AND PREDICTIVE MODELING OF CLIMATE CHANGE IN THE MOSCOW REGION USING MACHINE LEARNING METHODS

**I. Esipov
I. Staroverov**

Summary. The purpose of the study is to comprehensively assess the transformation of the climate regime in the Moscow region over an 86-year period (1940–2025) and to develop a long-term forecast of temperature trends up to the year 2040.

The research hypothesis is based on the assumption that regional warming has a pronounced asymmetric character with an emphasis on the winter period, and the application of additive machine learning algorithms allows for forecasting these climatic shifts more accurately than classical statistical methods.

To confirm the hypothesis, the following tasks were set: to conduct a retrospective analysis of meteorological time series, to quantitatively assess the shift of climatic seasons, and to perform comparative testing of the predictive ability of various ML models.

Research methods include the analysis of the ERA5 global atmospheric reanalysis dataset, the application of the non-parametric Mann-Kendall statistical test, and Sen's slope estimator. Predictive time series modeling was carried out using SARIMA, Prophet, and recurrent neural network (LSTM) models.

Results: Statistically significant asymmetric warming was revealed: winter temperatures rise 3.5 times faster than summer ones ($+0.535^{\circ}\text{C}$ vs. $+0.153^{\circ}\text{C}$ per decade). A shift in the onset date of climatic spring and a substantial decrease in the duration of snow cover were recorded. A comparative analysis of the algorithms showed that additive models (Prophet) are the most effective for long-term temperature forecasting. The constructed ensemble forecast indicates the persistence of the upward temperature trend and an expected increase in the mean annual temperature by $+0.77^{\circ}\text{C}$ by 2040 compared to 2025.

Keywords: climate change, Moscow region, global warming, Mann-Kendall test, time series analysis, machine learning, SARIMA, Prophet, LSTM.

Есипов Иван Владимирович

старший преподаватель, ФГБОУ ВО «МИРЭА,
Российский технологический университет»
iesipov98@yandex.ru

Староверов Игорь Николаевич

старший преподаватель, ФГБОУ ВО «МИРЭА,
Российский технологический университет»
St.igornik@yandex.ru

Аннотация. Цель исследования заключается в комплексной оценке трансформации климата в Москве за 86-летний период (1940–2025 гг.) и построении долгосрочного прогноза температурных трендов до 2040 года.

Гипотеза исследования базируется на предположении о том, что региональное глобальное потепление носит выраженный асимметричный характер с акцентом на зимний период, а применение аддитивных алгоритмов машинного обучения позволяет прогнозировать данные климатические сдвиги точнее классических статистических методов.

Для подтверждения гипотезы были поставлены следующие задачи: провести ретроспективный анализ метеорологических рядов, количественно оценить смещение климатических сезонов и выполнить сравнительное тестирование предсказательной способности различных ML-моделей.

Методы исследования включают анализ массива данных глобального атмосферного реанализа ERA5, применение непараметрического статистического критерия Манна-Кендалла и оценки уклона Сена. Предиктивное моделирование временных рядов осуществлялось с использованием моделей SARIMA, Prophet и рекуррентных нейронных сетей (LSTM).

Результаты: Выявлено статистически значимое асимметричное потепление: зимние температуры растут в 3,5 раза быстрее летних ($+0.535^{\circ}\text{C}$ против $+0.153^{\circ}\text{C}$ за десятилетие). Зафиксировано смещение даты наступления климатической весны и существенное сокращение продолжительности залегания снежного покрова. Сравнительный анализ алгоритмов показал, что для долгосрочного прогнозирования температур наиболее эффективны аддитивные модели (Prophet). Построенный ансамблевый прогноз указывает на сохранение восходящего температурного тренда и ожидаемый прирост среднегодовой температуры на $+0.77^{\circ}\text{C}$ к 2040 году по сравнению с 2025 годом.

Ключевые слова: изменение климата, Московский регион, глобальное потепление, тест Манна-Кендалла, анализ временных рядов, машинное обучение, SARIMA, Prophet, LSTM.

Введение

Глобальное изменение климата является одной из ключевых проблем XXI века. Согласно последним отчётам Межправительственной группы экспертов по изменению климата (МГЭИК / IPCC), рост глобальной приземной температуры носит беспрецедентный характер [1]. Однако на региональном уровне климатические изменения проявляются крайне неравномерно. Территория Российской Федерации, ввиду особенностей циркуляции атмосферы и эффекта «Арктического усиления» (Arctic amplification), нагревается в среднем в 2,5 раза быстрее, чем планета в целом [2, 3, 12].

Особый интерес для климатологии представляют крупные мегаполисы, такие как Москва, где на фоне глобальных атмосферных процессов действует локальный антропогенный фактор — эффект «городского острова тепла» (Urban Heat Island, UHI) [4]. Совместное влияние этих факторов приводит к трансформации температурного режима, изменению количества осадков и сдвигу климатических сезонов.

Цель данного исследования — на основе непрерывного 86-летнего массива данных провести строгую статистическую оценку трансформации климата Москвы, проанализировать динамику экстремальных погодных явлений и построить предиктивную модель температурного режима до 2040 года с применением алгоритмов машинного обучения.

Материалы и методы

В качестве источника использованы исторические метеорологические данные глобального атмосферного реанализа пятого поколения ERA5, предоставляемые

Европейским центром среднесрочных прогнозов погоды (ECMWF) через интерфейс Open-Meteo [5]. Исходный массив содержал почасовые наблюдения за период с 1 января 1940 г. по 31 декабря 2025 г. по координатам центра Москвы. Объем обработанных данных составил более 750 000 строк.

Предварительная обработка данных включала анализ пропусков и их линейную интерполяцию с учётом временного индекса. Периоды с полным отсутствием данных (например, глубина снежного покрова до 1950 года) были исключены из соответствующих расчётов.

Оценка качества исходного набора данных продемонстрировала высокую степень их полноты. Как видно на Рисунке 1, большинство параметров имеют нулевой процент пропусков. Наличие 100 % пропусков по показателям видимости (visibility) и эвапотранспирации (evapotranspiration) обусловлено отсутствием их фиксации в ранних архивах ERA5, в связи с чем данные признаки были исключены из анализа. Тепловая карта (Рисунок 2) позволяет определить временную локализацию пропущенных значений: в частности, данные о глубине снежного покрова отсутствуют исключительно до 1950 года, что было учтено при дальнейшей статистической обработке.

Для выявления монотонных тенденций во временных рядах применялся непараметрический статистический критерий Манна-Кендалла [13], устойчивый к выбросам. Величина тренда (градус/десятилетие) вычислялась с помощью непараметрической оценки уклона Сена [6]. Уровень статистической значимости был принят за $p < 0.05$.

Для предиктивного моделирования использовались три алгоритма различной математической природы:

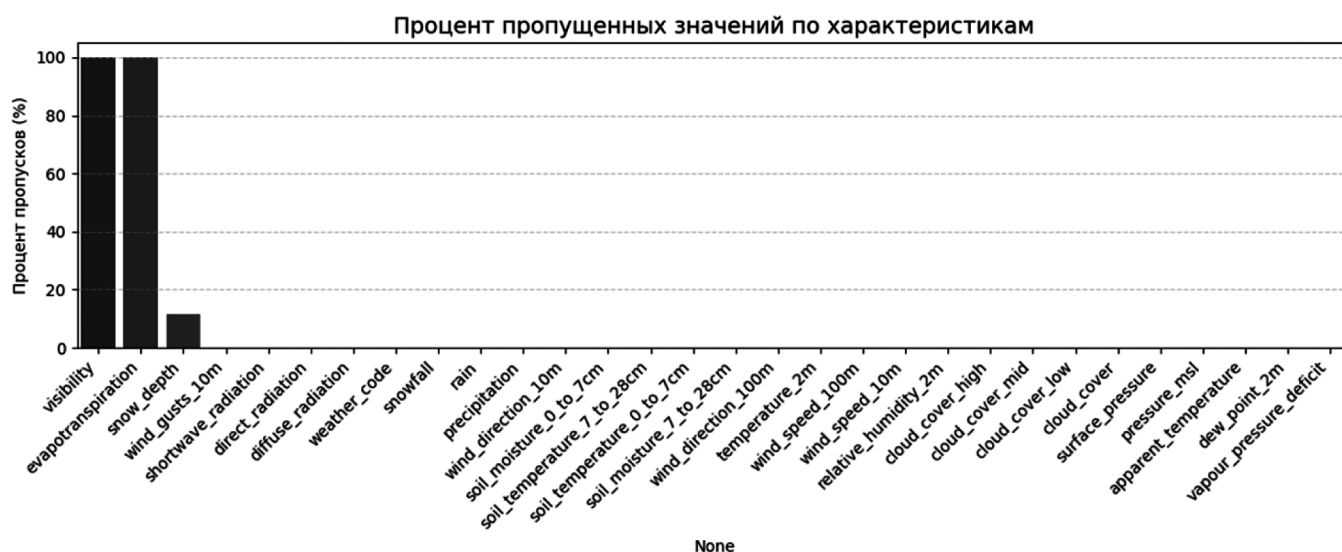


Рис. 1. Проценты пропущенных значений по каждому метеорологическому параметру в исходном наборе данных до интерполяции



Рис. 2. Тепловая карта распределения пропусков в данных по годам наблюдений

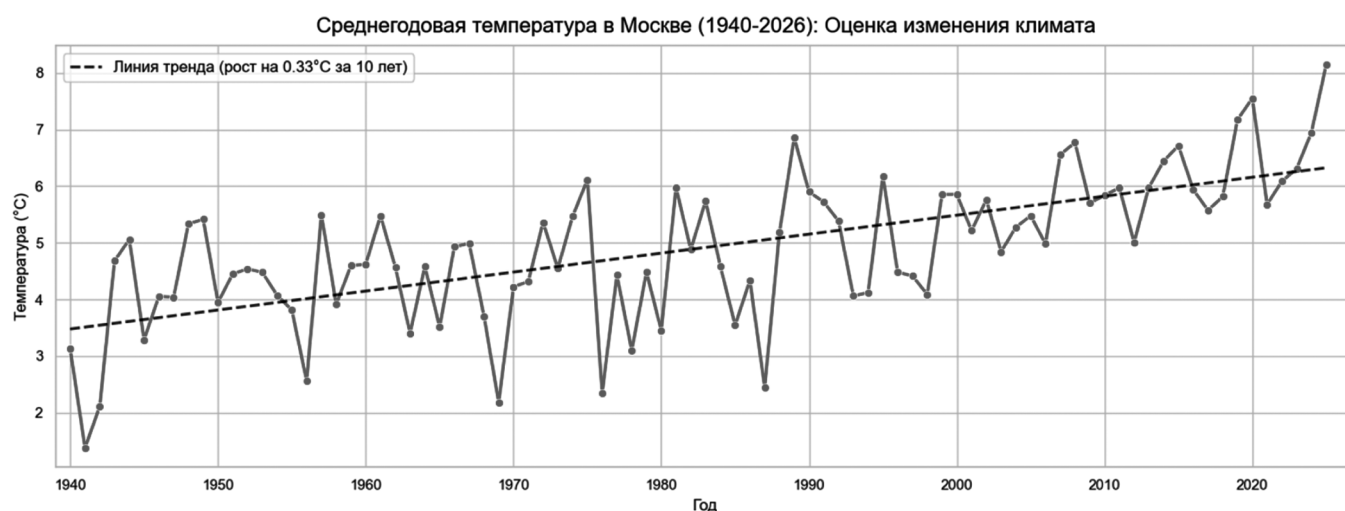


Рис. 3. Динамика среднегодовой температуры

1. Классическая статистическая модель **SARIMA** (Seasonal AutoRegressive Integrated Moving Average). [10]
2. Обобщённая аддитивная модель **Prophet**, разработанная для временных рядов с выраженной сезонностью [7].
3. Рекуррентная нейронная сеть долгой краткосрочной памяти **LSTM** (Long Short-Term Memory) [8].

Оценка качества моделей проводилась на отложенной тестовой выборке (2015–2025 гг.) с использованием метрик средней абсолютной ошибки (MAE) и корня из среднеквадратичной ошибки (RMSE).

Результаты

1. Глобальный тренд и асимметричность потепления

Анализ среднегодовых температур в Москве с 1940 по 2025 год выявил выраженный восходящий тренд.

Оценка Сена показала рост среднегодовой температуры со скоростью: $+0.323\text{ }^{\circ}\text{C}$ за 10 лет (статистическая значимость $p = 8.58 \times 10^{-11}$).

А динамика среднегодовой температуры на Рисунке 3 наглядно иллюстрирует непрерывный процесс глобального потепления.

Также, благодаря линейной аппроксимации ряда с 1940 по 2025 года подтверждено устойчивое смещение климатической нормы вверх без признаков замедления.

Изучение распределения температур по отдельным месяцам с помощью диаграмм размаха (Рисунок 4) показывает важную закономерность. Именно на зимний период приходится максимальная дисперсия показателей и основная масса температурных экстремумов.

Подобная вариативность и высокая частота статистических выбросов служат типичными маркерами умеренно-континентального типа климата.

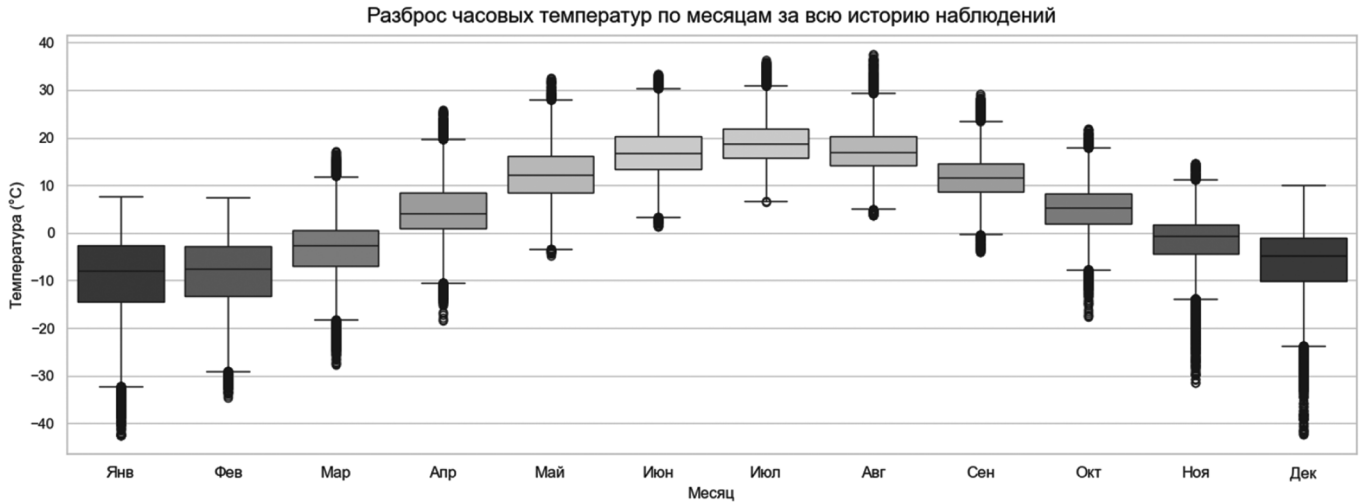


Рис. 4. Диаграмма размаха распределения часовых температур по месяцам года

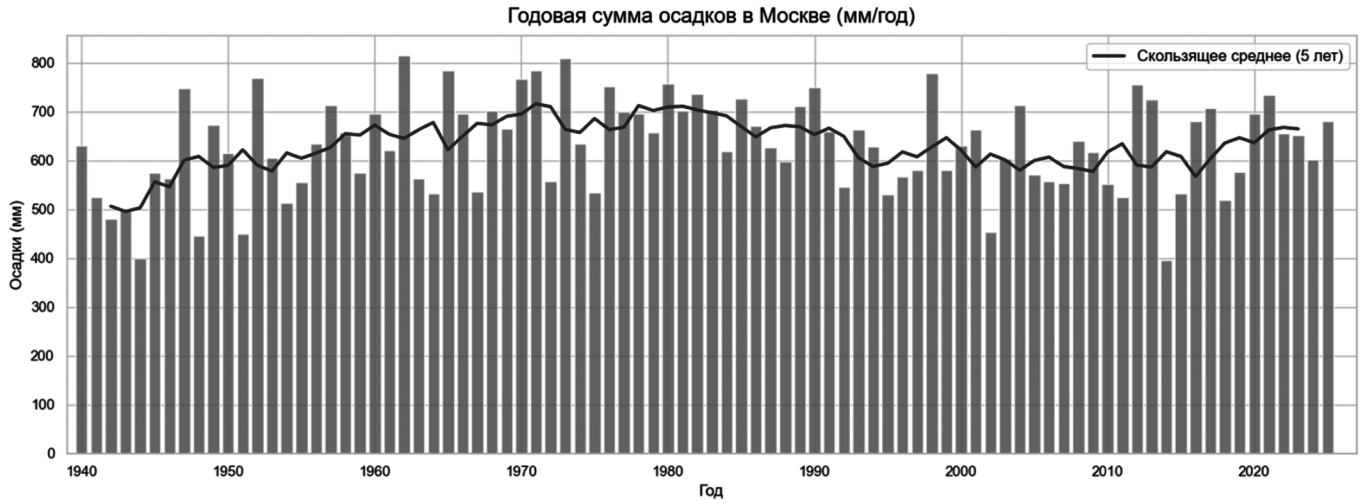


Рис. 5. Динамика годовой суммы осадков в Москве

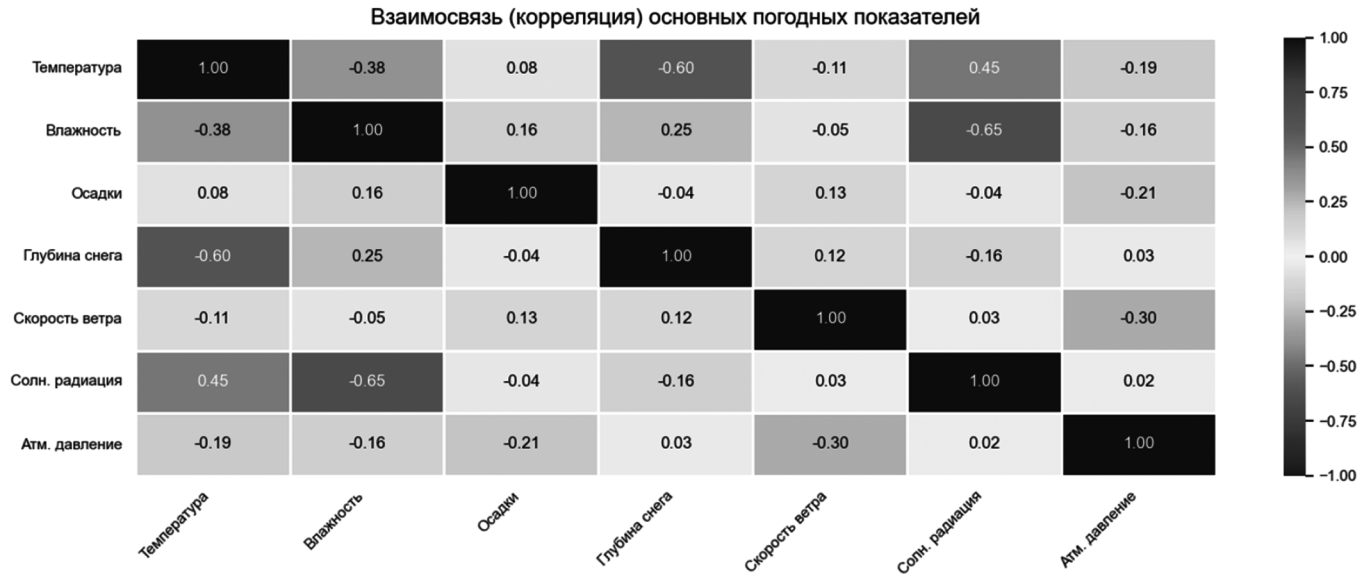


Рис. 6. Матрица парной корреляции Пирсона для основных климатических параметров

При анализе влажности (Рисунок 5) не было выявлено настолько сильных трендов, как при температуре. Пятилетнее скользящее среднее показывает циклические колебания суммарных годовых осадков в пределах 550–700 мм, характерные для данного региона.

Анализ полученной корреляционной матрицы (Рисунок 6) согласуется с известными физическими процессами в атмосфере. Как и следовало ожидать, рост температуры приводит к резкому сокращению снежного покрова, что даёт коэффициент обратной связи -0.60 . В свою очередь, увеличение солнечной радиации напрямую влияет на повышение температуры, показывая коэффициент: 0.45 .

Детализация данных по временам года выявила существенную внутригодовую неоднородность. Процесс регионального потепления идёт крайне неравномерно (Рисунок 7).

Больше всего заметно увеличение температуры в зимние месяцы, прирост составляет по $0.535\text{ }^{\circ}\text{C}$ за каждую декаду ($p=0.0001$). Весной прирост составляет около $0.368\text{ }^{\circ}\text{C}$ ($p<0.0001$), осенью: $+0.239\text{ }^{\circ}\text{C}$ ($p=0.0001$). Летом прирост лишь на $0.153\text{ }^{\circ}\text{C}$ за десять лет ($p=0.0083$).

Полученные значения доказывают, что зимы в Москве смягчаются в 3,5 раза быстрее, чем температура летом. То есть, основным источником изменения климата выступает именно исчезновение сильных морозов.

2. Динамика экстремальных явлений и смещение сезонов

Описанное выше отражается на частоте погодных аномалий. Например, после 2000 года наблюдается резкий обвал статистики по суровым морозным дням, когда столбик термометра опускается ниже $-20\text{ }^{\circ}\text{C}$. Параллельно с этим участились эпизоды летней жары с температурами от $30\text{ }^{\circ}\text{C}$ и выше.

Также, помимо самих экстремумов, заметно сдвинулись границы климатических сезонов. Дата прихода метеорологической весны (первый стабильный плюс после середины февраля) смещается на более ранние сроки: -1.8 дня за каждые 10 лет наблюдений ($p=0.008$). Оценка архива с 1950 года выявила строгую отрицательную динамику продолжительности снежного покрова ($p=0.0004$). Снег лежит в среднем на 3 дня меньше за каждое десятилетие.

Сокращение суровости зимнего периода наглядно отражено на Рисунке 9: если в середине XX века нормой считалось более 20–30 дней в году с температурой ниже $-20\text{ }^{\circ}\text{C}$, то в последние два десятилетия такие морозы фиксируются крайне редко. Одновременно с этим (Рису-

нок 8) возрастает частота появления дней с температурой свыше $30\text{ }^{\circ}\text{C}$, пик которых пришёлся на аномально жаркое лето 2010 года.

Последствием ускоренного потепления весеннего сезона является смещение фенофаз. Как видно на Рисунке 10, дата устойчивого перехода температуры через $0\text{ }^{\circ}\text{C}$ (начало климатической весны) постепенно сдвигается от начала апреля в сторону начала-середины марта.

Устойчивый тренд на сокращение продолжительности залегания снежного покрова подтверждается Рисунком 11. Линейный тренд иллюстрирует значительное падение количества снежных дней с 1950 по 2025 год.

3. Сравнительный анализ прогностических моделей (Machine Learning)

Для прогнозирования климатических изменений были обучены алгоритмы SARIMA, Prophet и LSTM на интервале 1940–2014 гг. Период 2015–2025 гг. использовался как тестовый. Результаты оценки метрик (Таблица 1) показывают, что классическая авторегрессионная модель (SARIMA) и аддитивная регрессия (Prophet) справились с задачей лучше, чем рекуррентная нейросеть (LSTM).

Таблица 1.
Сводная оценка метрик качества моделей
(Тестовый период 2015–2025)

Модель	MAE ($^{\circ}\text{C}$)	RMSE ($^{\circ}\text{C}$)
Prophet	1.967	2.586
SARIMA	1.973	2.560
LSTM	3.901	4.733

Визуальная оценка предсказательной способности моделей на тестовом участке 2015–2025 гг. демонстрирует, что классическая авторегрессионная модель SARIMA (Рисунок 12) и алгоритм Prophet (Рисунок 13) практически идеально улавливают сезонную амплитуду колебаний. Рекуррентная нейросеть LSTM (Рисунок 14) испытывает некоторые сложности с точным воспроизведением пиков экстремальных значений, что типично для сетей данного типа при ограниченном объёме данных.

Как легко заметить на итоговом сопоставлении (Рисунок 15), наилучшую сходимость с эмпирическими данными обеспечивает именно Prophet. Расчётная кривая данной модели максимально близко ложится на фактические значения. Закономерно, что столь высокая точность аппроксимации выразилась в минимальных значениях рассчитанных метрик (RMSE и MAE).

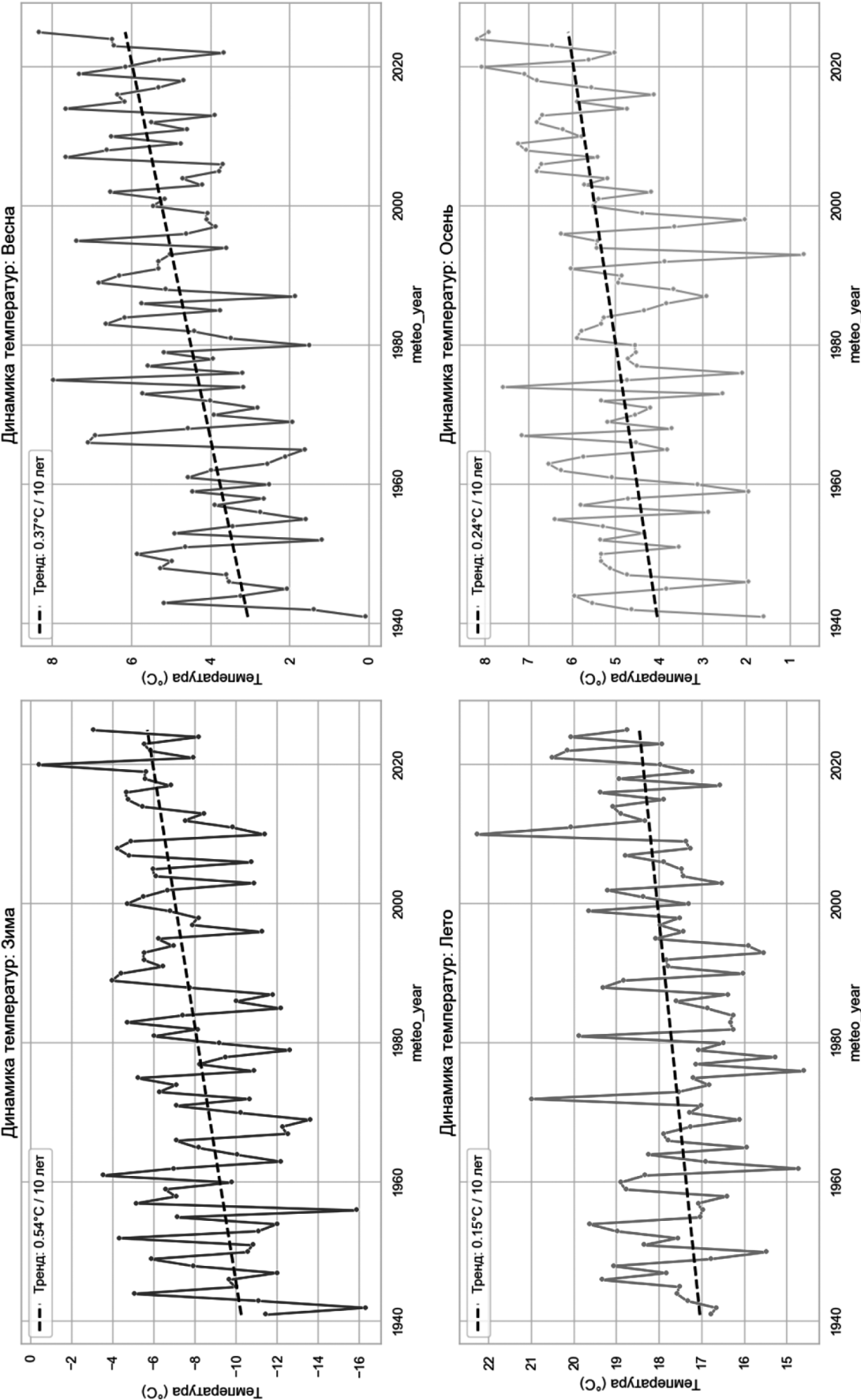


Рис. 7. Средние температуры по сезонам

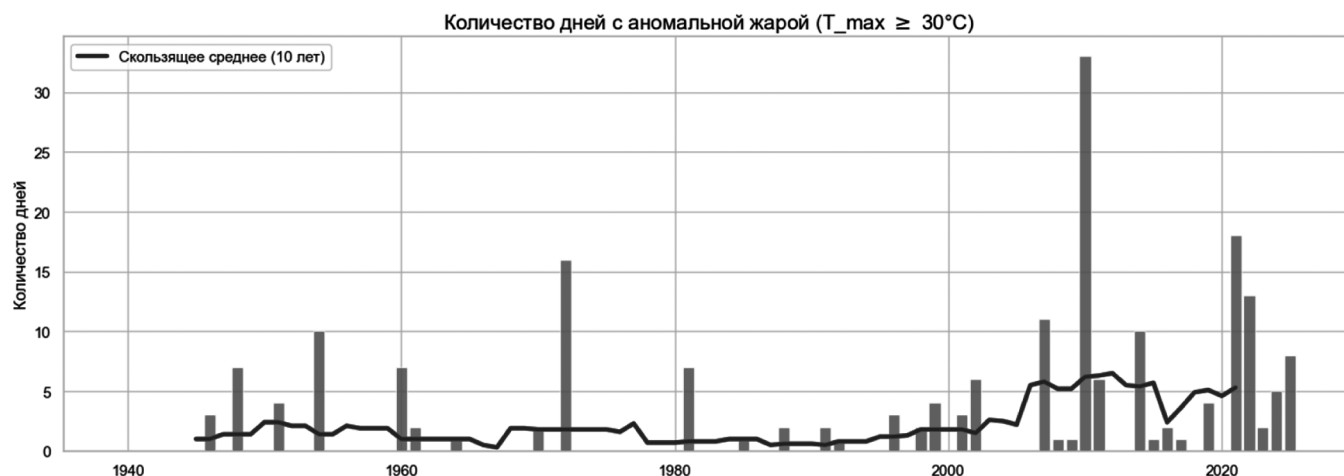
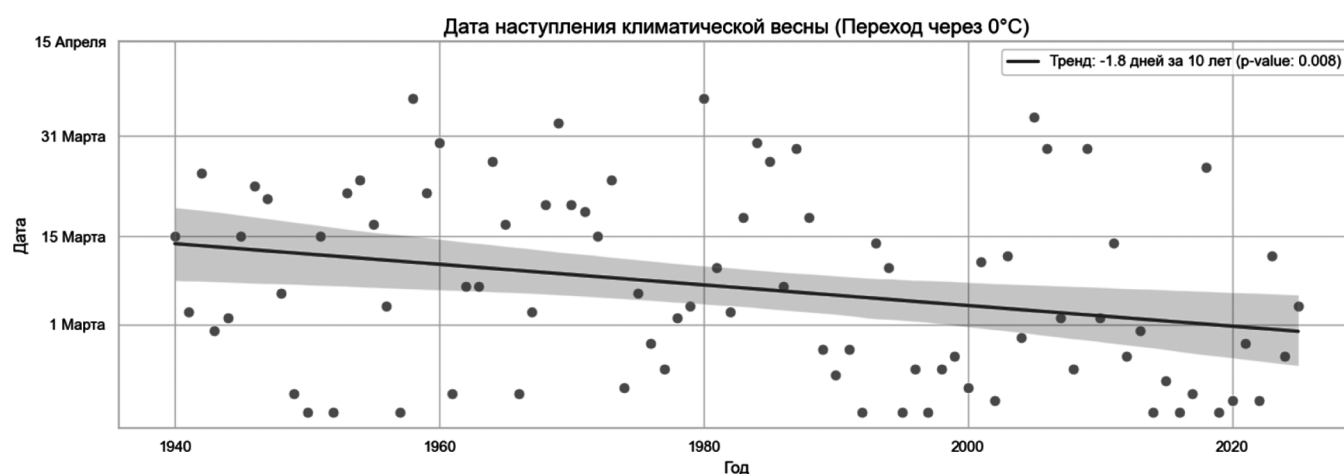


Рис. 8. Динамика количества дней в году с аномальной жарой



Рис. 9. Динамика количества дней в году с суровыми морозами

Рис. 10. Смещение даты устойчивого перехода среднесуточной температуры через 0°C (наступление климатической весны)

Относительно высокая погрешность рекуррентной модели LSTM связана с её архитектурными ограничениями. Классическое глубокое обучение довольно тяжело улавливает фиксированную годовую сезонность на стандартных метеорологических данных. Обычно

для повышения точности в таких задачах необходимо либократно увеличивать объем входных данных, либо переходить к комбинированным решениям формата CNN-LSTM.



Рис. 11. Изменение годового количества дней со стабильным снежным покровом (глубина > 1 см) в период с 1950 по 2025 г.



Рис. 12. Результаты прогнозирования среднемесячной температуры на тестовой выборке (2015–2025 гг.) с использованием модели SARIMA

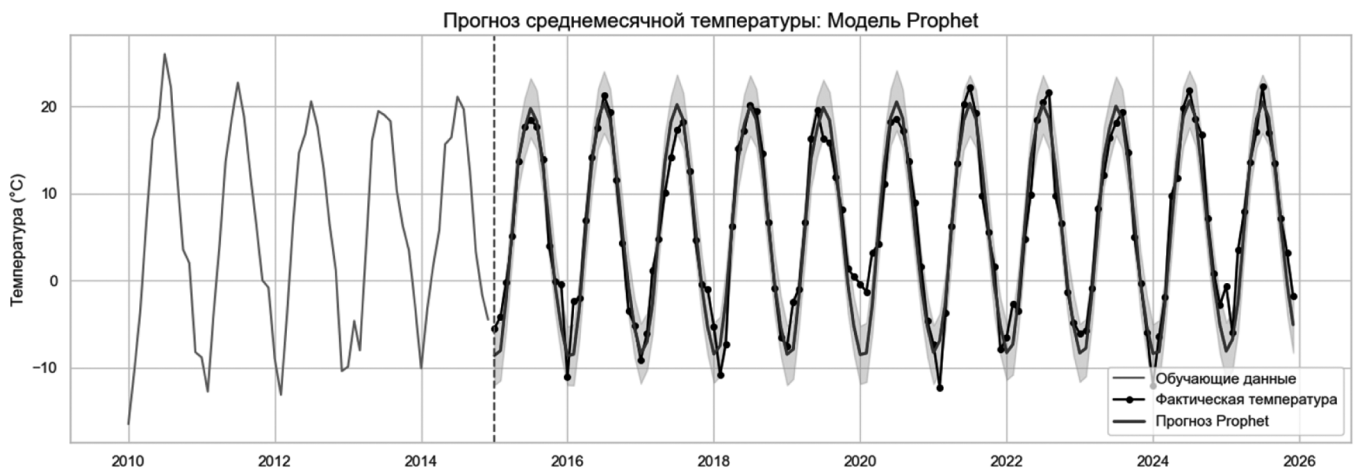


Рис. 13. Результаты прогнозирования среднемесячной температуры на тестовой выборке (2015–2025 гг.) с использованием Prophet

4. Прогноз

На основе полной выборки с 1940 по 2025 года был построен прогноз до 2040 года.

Итоговый ансамблевый прогноз (Рисунок 16) объединяет предсказания трех различных алгоритмов.



Рис. 14. Результаты прогнозирования среднемесячной температуры на тестовой выборке (2015–2025 гг.) с использованием LSTM

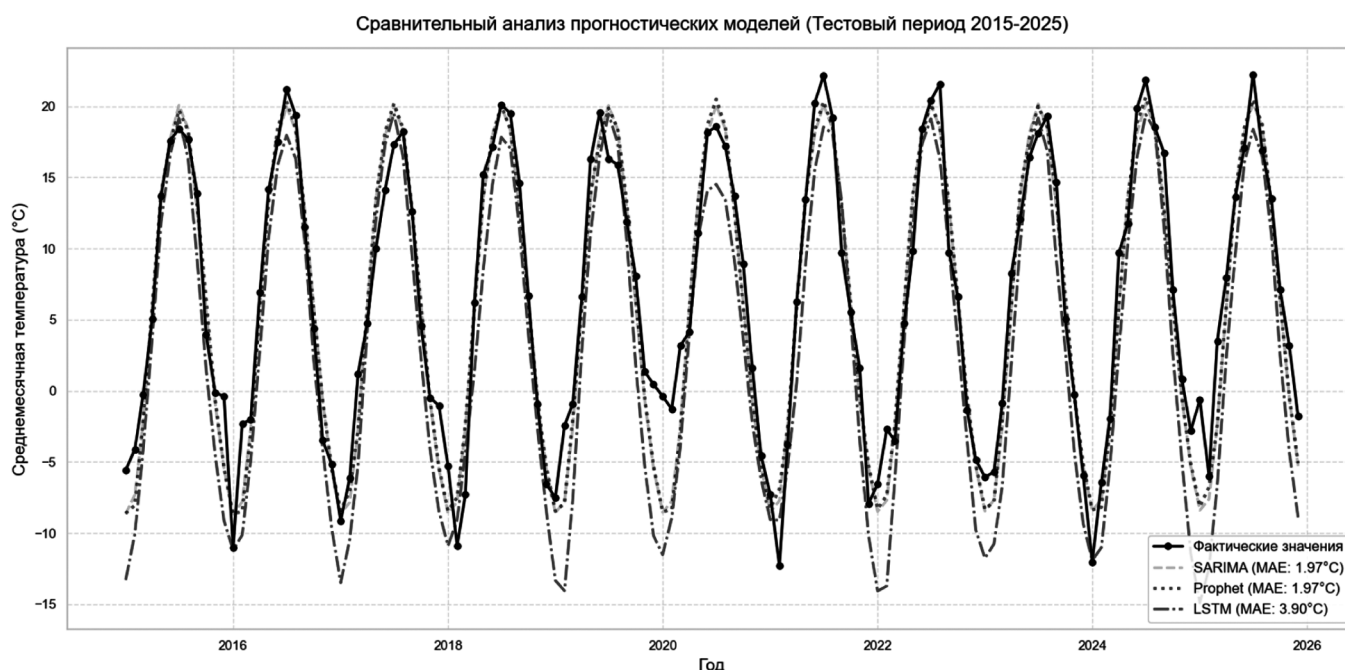


Рис. 15. Сравнение прогнозов моделей

Согласованность результатов всех трех моделей свидетельствует о высокой степени достоверности расчётов.

Базовая климатическая линия (тренд) модели Prophet демонстрирует устойчивый рост. При прогнозе мы видим, что базовая линия тренда возрастет с 6.51 °C в 2025 году, до 7.28 °C к 2040 году. Ожидаемый прирост среднегодовой температуры за следующие 15 лет составит: +0.77 °C.

Обсуждение результатов

Полученные результаты согласуются с выводами Росгидромета о том, что территория России подвержена беспрецедентному потеплению [3]. Обнаруженная нами

разница в скорости потепления зимы и лета физически обосновывается изменением макроциркуляционных процессов атмосферы: ослаблением влияния Сибирского антициклона и увеличением повторяемости выходов тёплых атлантических циклонов в зимний период [9].

Важным социально-экономическим последствием зафиксированного тренда (–3.0 дня снежного покрова за декаду и сокращение суровых морозов) является снижение нагрузки на отопительные системы мегаполиса. Однако выявленный в работе рост волн тепла ($T_{\max} \geq 30^{\circ}\text{C}$) требует адаптации городской инфраструктуры (системы кондиционирования, озеленение) для минимизации теплового стресса населения.

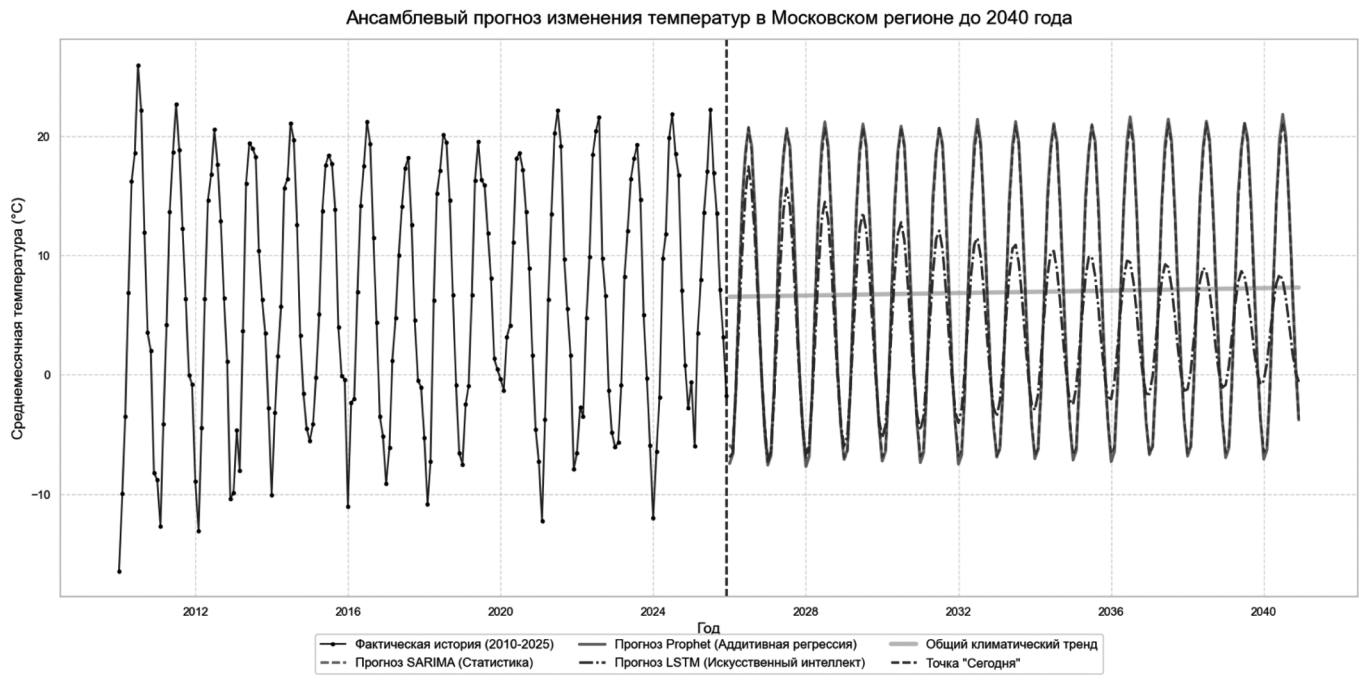


Рис. 16. Прогноз температуры до 2040 года с использованием всех моделей

Методологическим итогом работы стало подтверждение высокой стабильности (робастности) классических подходов к анализу временных рядов. Оказалось, что при прогнозировании атмосферных показателей на годы вперед именно SARIMA и Prophet лучше всего улавливают строгие сезонные циклы, не теряя при этом точности на длинной дистанции. Это подтверждено международными исследованиями [11].

Заключение

В результате анализа метеорологического архива Москвы за 86 лет, можно выделить несколько ключевых итогов.

Зафиксирован рост среднегодовой температуры, порядка $0.32\text{ }^{\circ}\text{C}$ за каждое десятилетие. Этот показатель

существенно обгоняет среднюю динамику изменения климата.

Доказана сезонная диспропорция. В зимние месяцы температура поднимается со скоростью $0.535\text{ }^{\circ}\text{C}$ за декаду, обгоняя летние темпы потепления в 3,5 раза. Это выразилось в том, что за 75-летний отрезок времени снежный покров стал лежать на 22 дня меньше.

Применение алгоритмов машинного обучения показало отличную предсказательную способность. Например, алгоритм Prophet, где средняя абсолютная ошибка составила $1.96\text{ }^{\circ}\text{C}$. Опираясь на эту модель, мы прогнозируем дальнейшее развитие теплового тренда и расчёты показали, что к 2040 году среднегодовая температура увеличится ещё на $0.77\text{ }^{\circ}\text{C}$.

ЛИТЕРАТУРА

1. IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
2. Доклад об особенностях климата на территории Российской Федерации за 2022 год. — М.: Росгидромет, 2023. — 104 с.
3. Катцов В.М., Порфирьев Б.Н. Оценка макроэкономических последствий изменений климата на территории Российской Федерации // Метеорология и гидрология. — 2011. — № 3. — С. 5–15.
4. Варенцов М.И., Константинов П.И., Бакланов А.А. и др. Исследование городского острова тепла Московского мегаполиса // Известия РАН. Физика атмосферы и океана. — 2017. — Т. 53, № 6. — С. 732–745.
5. Hersbach H., Bell B., Berrisford P. et al. The ERA5 global reanalysis // Quarterly Journal of the Royal Meteorological Society. — 2020. — Vol. 146, No. 730. — P. 1999–2049.
6. Sen P.K. Estimates of the Regression Coefficient Based on Kendall's Tau // Journal of the American Statistical Association. — 1968. — Vol. 63, No. 324. — P. 1379–1389.
7. Taylor S.J., Letham B. Forecasting at Scale // The American Statistician. — 2018. — Vol. 72, No. 1. — P. 37–45.

8. Hochreiter S., Schmidhuber J. Long Short-Term Memory // *Neural Computation*. — 1997. — Vol. 9, No. 8. — P. 1735–1780.
9. Бардин Ф.Ю., Полонский А.Б. Североатлантическое колебание и региональные изменения климата // *Известия РАН. Серия географическая*. — 2001. — № 2. — С. 16–27.
10. Box G.E.P., Jenkins G.M., Reinsel G.C., Ljung G.M. *Time Series Analysis: Forecasting and Control*. — 5th ed. — Hoboken, NJ: John Wiley & Sons, 2015. — 712 p.
11. Makridakis S., Spiliotis E., Assimakopoulos V. Statistical and Machine Learning forecasting methods: Concerns and ways forward // *PLoS ONE*. — 2018. — Vol. 13, No. 3. — e0194889.
12. Cohen J., Screen J.A., Furtado J.C. et al. Recent Arctic amplification and extreme mid-latitude weather // *Nature Geoscience*. — 2014. — Vol. 7, No. 9. — P. 627–637.
13. Mann H.B. Nonparametric tests against trend // *Econometrica*. — 1945. — Vol. 13, No. 3. — P. 245–259.

© Есипов Иван Владимирович (iesipov98@yandex.ru); Староверов Игорь Николаевич (Stigornik@yandex.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»

АЛГОРИТМ АВТОМАТИЧЕСКОГО ПОДДЕРЖАНИЯ РАССТОЯНИЯ ДО ОБРАБАТЫВАЕМОЙ ПОДЛОЖКИ В СИСТЕМАХ ЛАЗЕРНОЙ ОБРАБОТКИ¹

ALGORITHM FOR AUTOMATICALLY MAINTAINING THE DISTANCE TO THE PROCESSED SUBSTRATE IN LASER PROCESSING SYSTEMS

**D. Katasonov
V. Moskaleva**

Summary. A possible algorithm for automatically maintaining the distance between the working tool and the workpiece is considered using a complementary laser micromachining system as an example. The algorithm is based on automatic profilometry of the workpiece using a contactless distance sensor. The proposed approach can be applied to hybrid machining systems that perform layer-by-layer formation of objects. The results of a practical experiment are presented.

Keywords: laser micromachining, CAD model, automatic focusing, complementary system.

Катасонов Денис Николаевич

канд. тех. наук, старший научный сотрудник,
Федеральное государственное бюджетное учреждение
науки Институт автоматизации и электрометрии
Сибирского отделения Российской академии наук
(ИИАЭ СО РАН);

Доцент, Новосибирский государственный
технический университет (НГТУ)

katasonov@iae.nsk.su

Москалева Виолетта Михайловна

Новосибирский государственный технический
университет (НГТУ)

Аннотация. Рассматривается возможный вариант алгоритма автоматического поддержания расстояния между рабочим инструментом и обрабатываемой подложки на примере использования комплементарной системы лазерной микрообработки. В основе рассматриваемого алгоритма лежит автоматическая профилометрия обрабатываемой подложки при помощи бесконтактного датчика расстояния. Предложенный подход возможно применить в гибридных системах обработки, осуществляющих послойное формирование объектов. Показаны результаты практического эксперимента.

Ключевые слова: лазерная микрообработка, CAD-модель, автоматическая фокусировка, комплементарная система.

Введение

Осуществление лазерной обработки материалов подразумевает поддержание определенного расстояния между объективом, фокусирующим лазерный пучок и обрабатываемой подложкой. В зависимости от задачи — осуществление резки, микрообработки, гравировки и др. используются различные подходы для поддержания расстояния до обрабатываемой подложки. Одним из способов определения относительного положения подложки и объектива является использование бесконтактного датчика расстояния, на основе показаний которого осуществляется расчет положения. Кроме этого, похожая задача возникает при разработке гибридных систем формообразования [1]. Данное устройство осуществляет послойное нанесение пастообразной керамики на подложку, в соответствии с подготовленной CAD-моделью. В пределах области,

соответствующей нанесенному слою, осуществляется сканирование лазерного пучка, в результате чего происходит спекание нанесенного материала. В этом случае существует необходимость поддержания определенного расстояния поочередно для двух инструментов — диспенсера, который формирует слой и фокусирующего лазерный пучок объектива. Необходимость автоматического поддержания расстояния возникает, как при дефектах поверхности подложки, особенно в случае лазерной микрообработки, так и при лазерном и гибридном формообразовании на криволинейных подложках. Данная задача может быть решена путем формирования модели обрабатываемой поверхности в виде полинома, однако такой подход оказывается работоспособным в случае клиновидного дефекта подложки или неточного предварительного измерения толщины, выполняемой оператором.

¹ Работы выполнены в рамках проекта 124041700068-3 «Разработка новых технологий и аппаратных средств лазерной микрообработки и гибридного послойного формообразования с заданными фазовыми и структурными свойствами. Оптимизация характеристик лазерных гравиметров с использованием гибридных технологий»

В настоящей работе рассматривается алгоритм, основанный на предварительной бесконтактной профилометрии обрабатываемой подложки, при помощи триангуляционного датчика расстояния.

1. Устройство лазерной обработки

Апробация рассматриваемого в данной работе алгоритма осуществлялась на базе разработанной и выпускаемой ИАиЭ СО РАН комплементарной лазерной рабочей станции (ЛРС, complementary laser workstation, CLW) [2]. Рабочая станция состоит из объектива плоского поля и двух взаимодополняющих систем позиционирования луча — гальванометрических дефлекторов и шаговых приводов [3,4]. Схематичное изображение такой системы показано на рисунке Рис. 1.

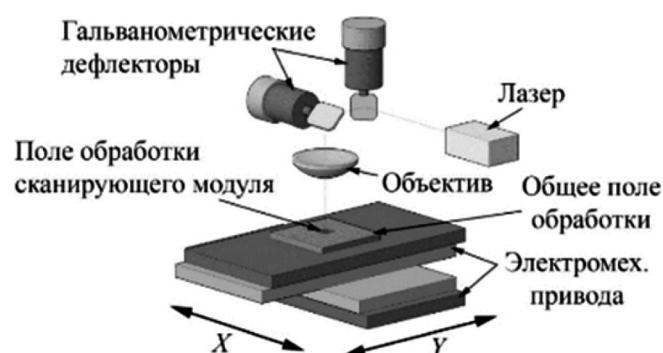


Рис. 1. Схема комплементарной системы лазерной обработки

Лазерный пучок попадает в расположенный над зоной обработки «низкоинерционный» сканер, который состоит из двух зеркальных гальванометрических дефлекторов. Фокусировка сканируемого пучка в пределах зоны обработки осуществляется объективом плоского поля. При этом, размеры и взаимное расположение формируемых объектов могут превышать размеры зоны обработки допустимые для фокусирующего объектива. В результате чего, обрабатываемая подложка располагается на предметном столе, который перемещается относительно фокусирующего объектива приводами «высокоинерционного» сканера. Низкоинерционный сканер и фокусирующий объектив выполнены в виде единого блока, который дополнительно содержит датчик расстояния, который будет использован для определения топологии подложки [5]. При этом, для обеспечения совмещения полей записи, на которые производится деление формируемого объекта, осуществляется коррекция дисторсии фокусирующего объектива [6,7]. Данный блок закреплен на вертикальном приводе, за счет которого осуществляется фокусировка лазерного пучка на подложке. Исходное положение вертикального привода определяется исходя из задаваемой оператором толщины подложки, а в процессе формирования объектов вносится коррекция в его положение на основе рассматриваемого в данной работе алгоритма.

Ранее был показан подход к осуществлению автоматической фокусировки примененных в рассмотренной системе микрообработки [8]. Такой подход подразумевает измерение расстояния до подложки непосредственно в пределах очередной зоны обработки. Измерение осуществлялось непосредственно перед формированием структуры в пределах этой зоны. Данный подход требовал значительного количества перемещений подложки между центром фокусирующего объектива и центром зоны измерения датчиком расстояния. Каждое перемещение требует технологической паузы необходимой для компенсации дрейфа положения высокоинерционного сканера, тем самым замедляется процесс формирования изделия. Также была показана возможность некоторой оптимизации количества таких перемещений путем решения транспортной задачи [9]. Данный подход разумно применять в случае формирования топологий на подложках, подверженных термическим деформациям, которые в свою очередь, приводят к изменению расстояния от подложки до фокусирующего объектива.

В данной работе рассматривается подход применимый для формирования топологий на подложках, не подверженных термическим деформациям, и подразумевает формирование траектории измерения подложки в пределах которых возможно осуществление непрерывных измерений (в пределах строки измерения, о чем будет сказано далее).

2. Построение карты высот подложки

Как было отмечено ранее, рассматриваемый алгоритм автоматической фокусировки заключается в предварительном измерении обрабатываемой подложки. В результате измерения формируется карта высот подложки в области в пределах которой будет осуществляться обработка. Для этого в виде прямоугольника определяется зона измерений на основе загруженной в СПО CAD-модели формируемого объекта. На основе полученного прямоугольника формируется сетка, в узлах которой осуществляется измерение. На рисунке 2 приведен пример подложки и формируемой топологии.

На рисунке 2 в качестве примера приведена наклонная плоскость, на которой формируется топология 1. При этом, запись осуществляется комплементарной системой, в результате чего топология делится на фрагменты 3. Автоматическая фокусировка осуществляется в два этапа — этапа подготовки и этапа коррекции. На предварительном этапе оператор задает расчетное значение толщины подложки, на основе которого формируется реперная координата положения вертикального привода. После чего автоматически осуществляется формирование сетки, в узлах 2 которой, показанных на рисунке 2, выполняются измерения. Сетка формируется на основе описывающего прямоугольника формируемой топо-

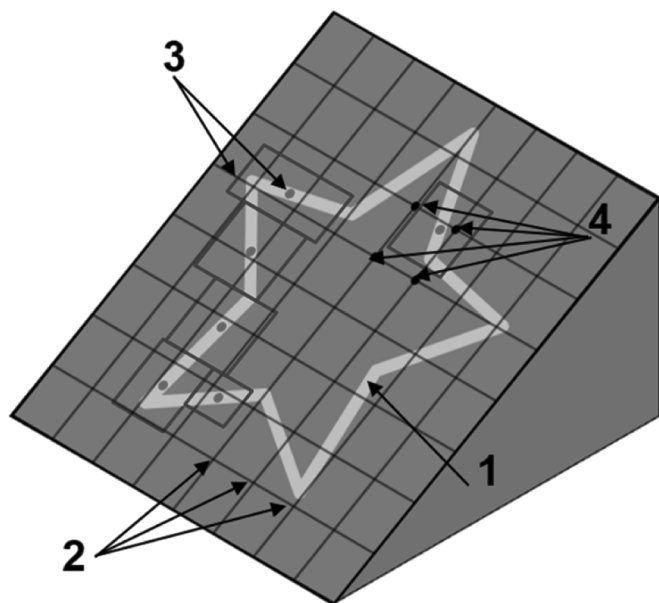


Рис. 2. Пример подложки и формируемой топологии.

1 — формируемая топология, 2 — узлы измерения, 3 — зоны лазерной записи, 4 — опорные точки

логии. Измерение осуществляется при непрерывном движении подложки при помощи высокоинерционного привода относительно измерительного датчика. Перед осуществлением измерений, вертикальный привод смещается от реперного положения на величину калибровочного смещения датчика расстояния. Калибровочное смещение определяется таким образом, чтобы показания датчика фокуса были нулевыми, при нулевой толщине заготовки. Таким образом, становится возможным определить величину отклонения толщины заготовки от расчетной или усредненной. На основе полученных измерений формируется карта высот обрабатываемой области. После того, как измерения выполнены, вертикальный привод перемещается в реперную координату.

Перед выполнением этапа коррекции осуществляется деление топологии 1 на участки обработки 3. Это вызвано особенностями комплементарной системы обработки. Далее, на этапе коррекции, для найденных ближайших точек 4 к центрам участков обработки 3 по выполненным измерениям осуществляется расчет величины коррекции положения вертикального привода от реперной позиции.

Алгоритм расчета величины коррекции

Узлы измерительной сетки не могут совпадать с центрами участков обработки. В результате чего, возникает необходимость поиска ближайших узлов сетки к центру очередного участка обработки. Компенсировать данную особенность можно выполнив измерения непосредственно в центрах участков обработки. Однако, такой подход оказывается неприменимым при осуществлении

последующего формирования заданной топологии. Это обусловлено тем, что положение участков обработки при формировании последующего слоя изменяется, как по причине связанной с самой топологией, так и с целью «размытия» границ между участками обработки. Таким образом, оказывается необходимым иметь изначальную карту высот подложки, и осуществлять расчет коррекции с ее помощью. Рассмотрим алгоритм поиска четырех ближайших точек. Поиск таких точек осуществляется путем выполнения следующих шагов:

1. Определяются координаты x , y центра очередного участка обработки.
2. Выполняется поиск четырех ближайших точек измерения. Для этого, в первую очередь, отсекаются все точки удаленные от заданной больше, чем на один шаг по x и по y . Если таких точек оказалось больше четырех выполняется сравнение весовых коэффициентов и определяются четыре ближайшие к заданной точке. Далее выполняется проверка, что выбранные точки формируют прямоугольник около заданной точки.
3. Производится расчет максимального расстояния max_x и max_y между измеренными отобранными ранее значениями.
4. Определяются расстояния dx и dy от заданной точки до каждой из четырех ближайших ранее найденных точек.
5. По полученным ранее значениям производится расчет весового коэффициента для каждой из четырех точек:

$$weight_x = max_x - dx$$

$$weight_y = max_y - dy$$

6. На основе полученных максимальных значений и весовых коэффициентов определяется значение величины коррекции для центра участка обработки:

$$h = \sum_{i=1}^4 \frac{h_i \cdot weight_x \cdot weight_y}{max_x \cdot max_y}$$

где h_i — высота (значение по z) для каждой i -той точки из четырех ближайших точек к заданной.

3. Результаты эксперимента

В качестве экспериментальной проверки предложенного алгоритма была использована подложка с намеренно введенной кривизной, на рисунке Рис. 3 показаны результаты профилометрии обрабатываемой подложки.

На рисунке 3 приведен фрагмент окна программного обеспечения сканирующего профилометра СКОПА-200 [10]. На рисунке показана карта высот 1, которая представлена в виде изображения, цвет каждого пикселя,

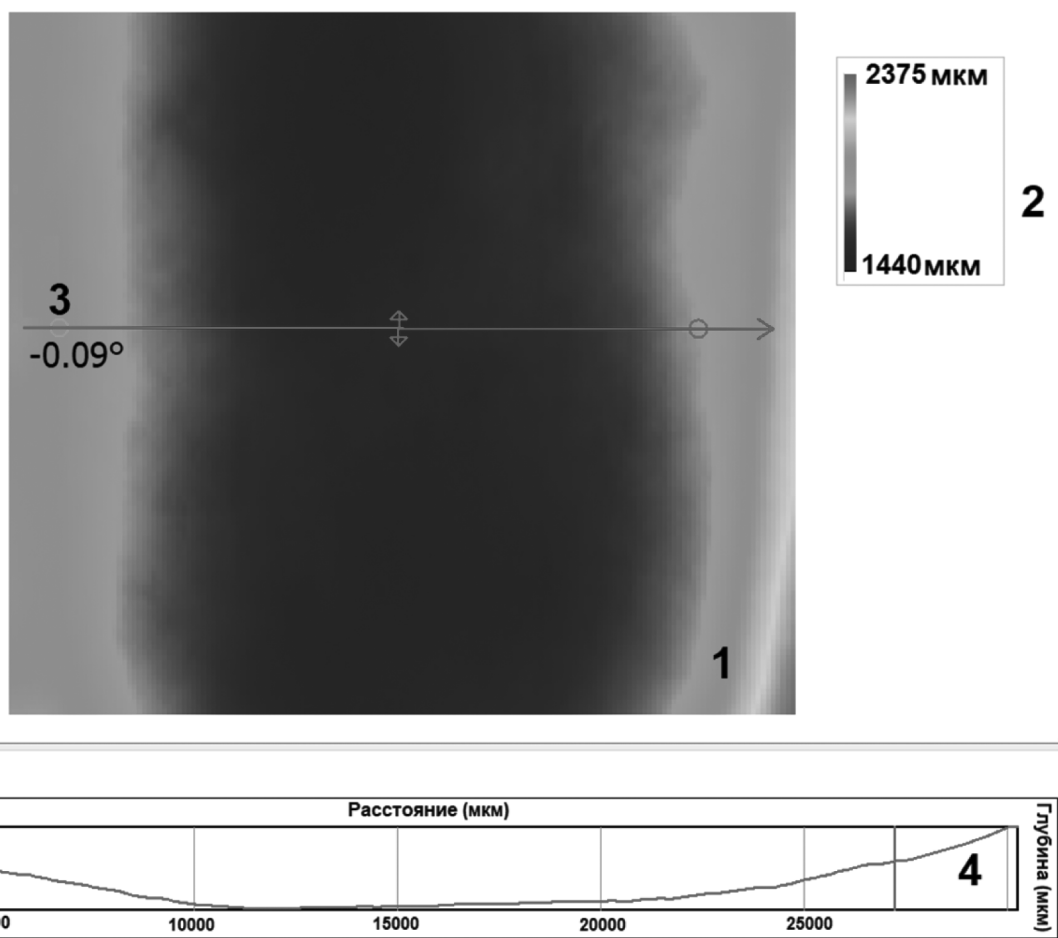


Рис. 3. Профилометрия обрабатываемой подложки.

1 — карта высот, 2 — палитра, 3 — линия, вдоль которой получен профиль, 4 — профиль

в соответствии с палитрой 2 кодирует высоту. Также на рисунке показан профиль 4, который сформирован вдоль линии 3. Как видно из построенного профиля, перепад высоты подложки находится в пределах 500 мкм, что превышает величину перетяжки лазерного луча.

В качестве эксперимента на подложке, профилограмма которой приведена на рисунке 4, была осуществлена лазерная запись структуры, состоящей из набора квадратов размером 2х2 мм, общий вид которых и стратегии их обработки приведен на рисунке 3.

На рисунке 4 приведен общий вид примитива, из которого составлена формируемая структура. Каждый примитив задается его внешним контуром, в пределах которых формируются траектории движения лазерного пучка. Траектории движения на рисунке 4 показаны в виде линий, которые располагаются под углом 45 градусов к горизонтالي с шагом в 100 мкм. Такой шаг сканирования выбран с целью наглядной визуализации нарушений фокусировки. Формируемая структура состоит из девяти таких примитивов, расположенных вдоль линий параллельных линии 3 на рисунке 3, в центральной части подложки. При этом, в качестве исходной толщи-

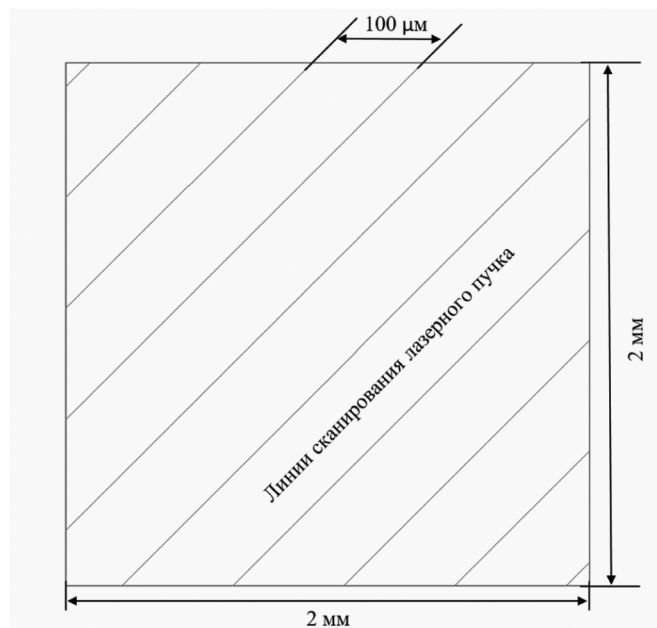


Рис. 4. Вид формируемого объекта и стратегия лазерной обработки

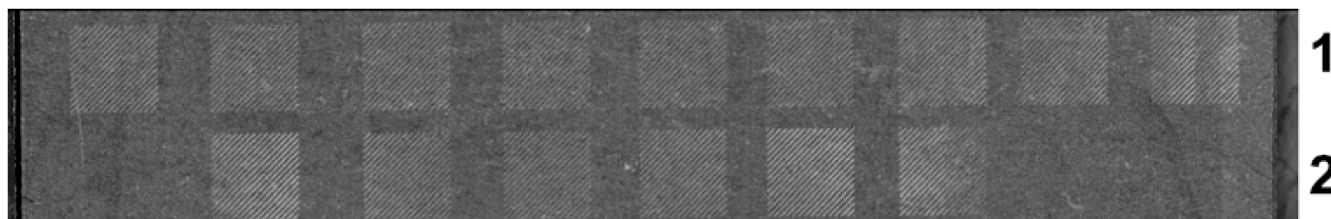


Рис. 5. Изображение сформированных объектов.

1 — структура сформированная с автоматической коррекцией, 2 — структура сформированная без коррекции

ны подложки было использовано значение, соответствующее центральной части подложки. При этом, поле записи было установлено равным 0.5x0.5 мм, для осуществления дополнительной проверки предложенного алгоритма, таким образом, один примитив формировался из четырех участков, для каждого из которых осуществлялась коррекция положения вертикального привода. На рисунке 5 приведен результат экспериментального формирования рассмотренной структуры на подложке, карта высот которой показана на рисунке 3.

Изображение сформированных объектов, приведенное на рисунке 5, было получено при помощи системы технического зрения профилометра СКОПА-200. Как

видно из фрагмента 2 на рисунке 5, на участках, соответствующих краям подложки запись не осуществляется, а структура приведенная на фрагменте 1 сформирована полностью.

Заключение

Рассмотрена возможная реализация алгоритма автоматической фокусировки лазерного пучка для комплементарной системы лазерной обработки, основанного на бесконтактной профилометрии обрабатываемой подложки при помощи триангуляционного датчика расстояния. Проведенные эксперименты показали работоспособность предложенного алгоритма.

ЛИТЕРАТУРА

1. Бессмельцев В.П., Голошевский Н.В., Катасонов Д.Н. Гибридный 3D-принтер для послойного формирования структур с проводящей топологией // Интерэкспо Гео-Сибирь. — 2021. — Т. 8. — С. 40–46.
2. Бессмельцев В.П., Баев С.Г., Булушев Е.Д., Голошевский Н.В., Горяев Е.П., Кастеров В.В., Максимов М.В., Смирнов К.К. Особенности технологии прецизионной фемтосекундной микрообработки оптических материалов // Интерэкспо Гео-Сибирь, № 2, 2016. с. 3–8.
3. Голошевский Н.В. Методы и программно-аппаратные средства управления устройствами лазерной микрообработки с комплементарной системой позиционирования: дис. канд. тех. наук. Новосибирск: Институт Автоматики и Электрометрии СО РАН, 2021. 138 с.
4. Goloshevsky N. et al. Precision laser system based on complementary scanning principle for dielectric materials microprocessing // Fundamentals of Laser Assisted Micro-and Nanotechnologies. — SPIE, 2008. — Т. 6985. — С. 170–178.
5. Бессмельцев В.П., Голошевский Н.В. Система управления составными двухкоординатными сканаторами // Автометрия. 2007. № 1. С. 116–126
6. Бессмельцев В.П., Голошевский Н.В., Кастеров В.В., Киприянов Я.А., Смирнов К.К. Метод калибровки лазерной гальванометрической сканирующей системы субмикронного разрешения // Автометрия, № 4, 2018. с. 84–92.
7. Weisz J.R. Software calibration of scan system distortions // Proc. SPIE. 1991. 1454. P. 265–271
8. Баев С.Г. и др. Применение алгоритмов автоматической фокусировки в процессе лазерной субмикронной микрообработки. // Системы анализа и обработки данных. — 2026. — Т. 2. — № 4. — С. 147–158.
9. Катасонов Д.Н., Москалева В.М. Оптимизация траекторий перемещения при автоматической фокусировке в комплементарных системах лазерной микрообработки // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и Технические Науки. — 2026. — №02. — С. 75–80 DOI 10.37882/2223-2966.2026.02.13
10. «ГСИ. Профилометры сканирующие конфокальные СКОПА-200. Методика поверки» МП 017.М44-25. // Москва. — 2025.

© Катасонов Денис Николаевич (katasonov@iae.nsk.su); Москалева Виолетта Михайловна
Журнал «Современная наука: актуальные проблемы теории и практики»

ФАКТОРЫ, СНИЖАЮЩИЕ РЕЗУЛЬТАТИВНОСТЬ УПРАВЛЕНИЯ УСТАРЕВАНИЕМ ОБОРУДОВАНИЯ В ОРГАНИЗАЦИЯХ

FACTORS REDUCING THE EFFECTIVENESS OF EQUIPMENT OBSOLESCENCE MANAGEMENT IN ORGANIZATIONS

**O. Klementyeva
O. Chmylevskaya**

Summary. The purpose of the authors' research is to develop a structured approach to managing equipment obsolescence in resource-intensive industries (oil and gas, mining, manufacturing), which allows converting the uncertainty associated with obsolescence into a controlled risk.

As a research approach, the authors analyzed the concept of equipment obsolescence from a practical point of view, highlighting three aspects: technical, functional and economic; the William Edwards Deming «Planning — Action — Verification—Adjustment» (PDCA) model was applied to build a structured approach to obsolescence management, and specific measures were developed for each stage. PDCA models.

Key results include a list of reasons for ineffective obsolescence management; detailed measures for each stage of the PDCA model, and a set of key performance indicators for monitoring the obsolescence management process.

The scientific significance of the work consists in expanding the theoretical foundations of equipment obsolescence management by introducing a three-component obsolescence model (technical-functional-economic aspects) and adapting the cyclic PDCA model to this issue; developing a methodology for proactive obsolescence management as an alternative to the reactive approach.

The practical significance of the author's work is to provide enterprises in resource-intensive industries with a structured approach to obsolescence management ready for implementation, as well as tools for assessing obsolescence management through the KPIs system.

Keywords: equipment obsolescence, risk management, Deming model (PDCA), key performance indicators (KPIs), asset lifecycle, proactive management, production reliability.

Клементьева Ольга Васильевна

ведущий эксперт технического контроля
АНО ДПО «Институт организационного развития
и лидерства»

Чмилевская Оксана Сергеевна

генеральный директор, АНО ДПО «Институт
организационного развития и лидерства»

Аннотация. Цель исследования авторов — разработка структурированного подхода к управлению устареванием оборудования в ресурсоёмких отраслях (нефтегазовая, горнодобывающая, обрабатывающая промышленность), позволяющий преобразовать неопределённость, связанную с устареванием, в контролируемый риск.

В качестве исследовательского подхода авторами был проведён анализ понятия устаревания оборудования с практической точки зрения с выделением трёх аспектов: технического, функционального и экономического; была применена модель Уильяма Эдвардса Деминга «Планирование — Действие — Проверка — Корректировка» (PDCA) для построения структурированного подхода к управлению устареванием, а также разработаны конкретные мероприятия для каждого этапа модели PDCA.

Ключевые результаты включают в себя перечень причин неэффективного управления устареванием; детализированные мероприятия по каждому этапу модели PDCA, набор ключевых показателей эффективности для контроля процесса управления устареванием.

Научная значимость работы состоит в расширении теоретических основ управления устареванием оборудования за счёт введения трёхкомпонентной модели устаревания (технический-функциональный-экономический аспекты) и адаптации циклической модели PDCA к данной проблематике; разработка методологии проактивного управления устареванием как альтернативы реактивному подходу.

Практическая значимость авторской работы состоит в предоставлении предприятиям ресурсоёмких отраслей готового к внедрению структурированного подхода к управлению устареванием, инструментария для оценки управления устареванием через систему КПЭ.

Ключевые слова: устаревание оборудования, управление рисками, модель Деминга (PDCA), ключевые показатели эффективности (КПЭ), жизненный цикл актива, проактивное управление, надёжность производства.

Введение

Устаревание оборудования на производстве редко проявляется в структурированном или управляемом виде, вместо этого оно проявляется в виде незапланированных простоев, увеличения сроков выполнения заказов, роста затрат на техническое обслуживание и принятия инженерных решений в последнюю минуту в условиях производственной необходимости. В таких ресурсоёмких отраслях как нефтегазовая, гор-

нодобывающая, и обрабатывающая промышленность, неспособность предвидеть устаревание и управлять им является не техническим ограничением, а системным сбоем в том, как организации сочетают мышление о жизненном цикле объекта, управление рисками и исполнительскую дисциплину.

На практике разница между организациями, которые борются с повторяющимися проявлениями устаревания, и теми, которые эффективно справляются с ними, за-

ключается в их способности рассматривать устаревание оборудования как предсказуемый и управляемый риск, а не как случайный сбой. Это требует комплексного подхода, который согласовывает функции проектирования, технического обслуживания, планирования и цепочки поставок в рамках структурированной системы, которая является практичной и устойчивой в условиях реальных ограничений, таких как ограниченные ресурсы, несовершенные данные и конкурирующие операционные приоритеты.

В условиях высокой капиталоемкости, длительного жизненного цикла активов и жестких требований к надёжности производственных систем, проблема устаревания оборудования приобретает характер не локального технического затруднения, а системного управленческого риска. На практике устаревание проявляется через рост вероятности незапланированных простоев, увеличение стоимости технического обслуживания, снижение доступности запасных частей, усложнение модернизации и усиление зависимости организации от срочных инженерных и закупочных решений. Особенно остро данная проблема проявляется в нефтегазовой, горнодобывающей и обрабатывающей промышленности, где цена отказа определяется не только прямыми затратами, но и рисками для безопасности и целостности объекта, исполнения договорных обязательств.

Обзор существующих исследований

Современные исследования показывают, что управление устареванием оборудования постепенно смещается от узкотехнического подхода к междофункциональной модели, учитывающей жизненный цикл актива, неопределённость поставок, организационные процессы, риски совместимости и экономическую эффективность решений.

1. Длительные контракты и жизненный цикл объекта

В работах [1, 2] предложены модели согласования условий долгосрочных контрактов с этапами жизненного цикла, однако основное внимание уделяется начальной фазе (проектирование, строительство) и не учитываются риски технологического устаревания на этапе эксплуатации. В нескольких исследованиях [6, 4] особое внимание уделяется системам активного планирования и раннему выявлению рисков поставок и по сути, долгосрочные соглашения должны учитывать, что структура затрат не является статичной на протяжении нескольких десятилетий, но это не связывается с показателями устаревания. В работах [1, 2, 8] неизменной темой является неопределённость, связанная с длительным сроком службы, обусловленная развитием технологий, изменениями в цепочках поставок и меняющимися экосистемами поддержки.

Таким образом, актуальной остаётся задача разработки контрактных механизмов, которые бы динамически адаптировались к риску устаревания оборудования на всех этапах жизненного цикла объекта, обеспечивая баланс интересов заказчика и поставщика.

2. Риск-ориентированный подход

В исследованиях [1,2] моральное устаревание последовательно рассматривается как область многомерных рисков, включающая техническую, операционную, экономическую неопределённость и неопределённость цепочки поставок. В прикладном исследовании [5,7] представлены некоторые структурированные методологии, но подчеркиваются практические трудности при проведении оценки рисков, включая ограниченность данных и несогласованность критериев оценки. Исследования различаются по тому, как они определяют и классифицируют риски: [9] техническое и функциональное устаревание, [10] доступность цепочки поставок и поддержка поставщиков, шкалы оценки рисков различаются (качественные, полуколичественные и количественные показатели), пороговые значения для «высокого риска» противоречивы, терминология значительно различается, что ограничивает сравнительный анализ, обмен данными и внедрение передовых практик между операторами и регионами. Анализ литературы показывает, что достигнута зрелость в принятии основанных на оценке рисков подходов к управлению устареванием, но недостатком является отсутствие стандартизированной интегрированной системы классификации рисков, способной поддерживать последовательный процесс принятия решений по долгосрочным и сложным активам. До тех пор, пока отрасль не придет к единому классификационному подходу, организации будут продолжать эффективно справляться с устареванием на местном уровне, но будут бороться за его оптимизацию в масштабах системы и предприятия.

3. Отраслевые решения для нефтегазовой и тяжелой промышленности

Анализ источников ссылок показывает, что методы управления устареванием больше не являются теоретическими — они активно внедряются в нефтегазовой и тяжелой промышленности на международной арене.

Исследования [3, 4], подтверждают, что ведущие операторы интегрируют устаревание оборудования в долгосрочные стратегии продления срока службы активов, планируют жизненный цикл объекта, встраивая вопросы устаревания в стратегии техобслуживания. Также согласно анализу источников [9,10] происходит признание взаимозависимости между оборудованием, программным обеспечением и поставщиками, формируются соглашения с поставщиками о расширенной поддержке.

Однако степень зрелости методов неодинакова, с очевидными преимуществами в структурированных подходах и столь же очевидными недостатками в интеграции, масштабируемости и последовательности исполнения, многие организации сосредоточены на отдельных компонентах или оборудования, а не на системах и испытывают трудности с преобразованием рисков в финансовые последствия.

4. Специализированные Ключевые Показатели Эффективности (КПЭ)

В основополагающих и недавних работах [1, 2] эффективность управления устареванием, как правило, оценивается с помощью комбинации показателей эффективности эксплуатации, рисков и жизненного цикла. Также в промышленности широко используются [6] объемы невыполненной работы, процент замененных или устаревших деталей, экстренные мероприятия, связанные с устареванием. При этом существуют концепции [6], где разработаны индикаторы, основанные на оценке рисков, такие как количество элементов, связанных с высоким риском устаревания, снижение критического риска со временем и проч. При этом, при наличии диапазона показателей, выявляется чрезмерная зависимость от реактивных показателей, таких как учет выявленных проблем, предпринятых действий и количество сбоев.

До тех пор, пока ключевые показатели эффективности не перейдут от отслеживания деятельности к демонстрации ценности и поддержке принятия решений, пока будет наблюдаться отсутствие общих определений КПЭ в разных организациях, разные значения для высокого риска и не стандартизированные методы расчета, управление устареванием будет оставаться заметным в оперативном плане, но стратегически недостаточно эффективным.

Проведенный анализ научных исследований и отраслевых практик показывает, что проблема управления устареванием оборудования в нефтегазовой отрасли требует комплексного подхода, объединяющего:

- Учёт долгосрочных контрактов и этапов жизненного цикла;
- Применение риск-ориентированного мышления для прогнозирования;
- Интеграцию отраслевых решений в единую систему;
- Разработку специализированных КПЭ, отражающих как технические, так и контрактные риски.

Существующие модели и методики решают отдельные аспекты этой задачи, но не обеспечивают её целостного решения.

В следующем разделе представлена авторская модель, которая устраняет выявленные пробелы за счёт

связывания классификации рисков устаревания, причин организационных сбоев, цикла управленческих действий и систему показателей результативности в едином контуре принятия решений.

Анализ текущей практики показывает, что во многих организациях управление устареванием остаётся реактивным: признаки потери ремонтпригодности, прекращения поддержки, снижения совместимости или экономической целесообразности фиксируются слишком поздно, а решения принимаются уже в условиях дефицита времени, данных и доступных альтернатив. Такая ситуация указывает на необходимость перехода от фрагментарного реагирования к системному риск-ориентированному управлению, встроенному в процессы эксплуатации, технического обслуживания, снабжения и планирования.

В связи с этим представляется актуальной разработка концептуальной модели управления устареванием оборудования, которая позволяла бы рассматривать устаревание не как единичное событие, а как предсказуемый и контролируемый риск, требующий регулярной идентификации, оценки, мониторинга и корректирующих действий. Такой подход может стать методической основой для повышения надёжности производственных систем, снижения операционных потерь и повышения обоснованности управленческих решений в условиях технологической и рыночной неопределённости.

Цель исследования — разработать риск-ориентированную концептуальную модель управления устареванием оборудования в ресурсоёмких отраслях, объединяющую классификацию видов устаревания, анализ причин организационной неэффективности, цикл PDCA и систему ключевых показателей эффективности для мониторинга и снижения риска.

Для достижения поставленной цели решались следующие задачи:

1. Проанализировать существующие подходы к пониманию устаревания оборудования в научной и нормативной литературе.
2. Систематизировать основные проявления устаревания оборудования в производственной среде.
3. Выделить и описать ключевые аспекты устаревания: технический, функциональный и экономический.
4. Определить типичные организационные причины неэффективного управления устареванием оборудования.
5. Разработать риск-ориентированную модель управления устареванием на основе цикла Шухарта-Деминга «PDCA».
6. Предложить систему КПЭ, позволяющую оценивать уровень готовности организации к рискам

устаревания и результативность принимаемых решений.

7. Сформулировать практические ориентиры для внедрения системы управления устареванием в деятельность промышленных организаций.

Объект исследования — процессы управления жизненным циклом оборудования в организациях ресурсёмких отраслей промышленности.

Предмет исследования — организационно-управленческие механизмы выявления, оценки, мониторинга и снижения рисков, связанных с устареванием оборудования.

Гипотеза исследования состоит в том, что эффективность управления устареванием оборудования существенно возрастает, если рассматривать устаревание не как локальную техническую проблему, а как системный риск и управлять им на основе интеграции:

1. классификации видов устаревания;
2. межфункционального процесса принятия решений;
3. цикла PDCA;
4. системы проактивных и реактивных КПЭ.

При таком подходе возможно повысить предсказуемость отказов, снизить долю реактивных решений и улучшить экономические и операционные показатели эксплуатации оборудования.

Методологическую основу исследования составили общенаучные и специальные методы. В работе использованы: анализ и синтез научной литературы по проблеме управления устареванием оборудования; сравнительный анализ нормативных и методических документов; классификация видов устаревания; структурно-функциональный анализ причин неэффективного управления; процессное моделирование управленческого цикла на основе PDCA; метод концептуального проектирования системы КПЭ для мониторинга состояния и результативности управления устареванием.

Научная новизна исследования заключается в следующем:

1. Устаревание оборудования обосновано как системный управленческий риск, а не только как техническое следствие физического или морального износа.
2. Предложена трёхкомпонентная классификация устаревания оборудования, включающая технический, функциональный и экономический аспекты, рассматриваемые во взаимосвязи.
3. Систематизированы организационные причины неэффективного управления устареванием, включая реактивный стиль управления, недостаточ-

ную полноту данных, размытость ответственности и преобладание реактивных показателей.

4. Разработана концептуальная риск-ориентированная модель управления устареванием оборудования на основе цикла PDCA, интегрирующая идентификацию риска, принятие решений, реализацию мер, мониторинг и корректирующие действия.
5. Предложена система КПЭ для оценки состояния и результативности управления устареванием, включающая как проактивные, так и реактивные показатели.

Теоретическая значимость исследования состоит в развитии представлений об устаревании оборудования как междисциплинарной категории на стыке управления рисками, надёжности, технического обслуживания и организационного проектирования. Предложенная модель расширяет существующие подходы к управлению устареванием оборудования за счёт объединения классификации рисков, процессной логики управления и системы показателей в единой аналитической рамке.

Практическая значимость исследования заключается в возможности использования предложенной модели промышленными предприятиями для: раннего выявления компонентов и систем, подверженных риску устаревания; распределения ответственности между подразделениями; планирования мероприятий по замене, модернизации или закупке; построения системы мониторинга на основе КПЭ; снижения доли аварийных и срочных решений; повышения предсказуемости затрат и устойчивости производственных процессов.

Мы полагаем, что устаревание оборудования должно рассматриваться не как эпизодическая проблема снабжения или ремонта, а как неотъемлемая часть жизненного цикла актива, требующая встроенного межфункционального управления и регулярного мониторинга на основе проактивных и реактивных показателей.

Определение устаревания с практической точки зрения

В реальных производственных условиях устаревание редко фиксируется как формальное условие в системах, но вместо этого проявляется в процессе эксплуатации из-за увеличения трудностей при проведении технического обслуживания. Бригады технического обслуживания сталкиваются с такими фактами как: детали, которые больше не поставляются производителями оригинального оборудования, системы управления, которые технически исправны, но больше не поддерживаются, и программные среды, которые невозможно обновить без риска нестабильности системы. Параллельно орга-

низации сталкиваются с консолидацией поставщиков или выходом из определенных продуктовых линеек, в то время как внутри компании постепенная потеря опытного персонала приводит к устареванию навыков, на которое часто не обращают внимания, пока оно не становится критическим.

С практической точки зрения, риски устаревания можно рассматривать в трех взаимосвязанных аспектах.

Техническое устаревание происходит, когда компоненты или системы больше не обслуживаются или не производятся, что создает непосредственный риск для поставок, некорректно оценивается критичность оборудования, отсутствует поддержка программного обеспечения.

Функциональное устаревание возникает, когда активы больше не могут соответствовать эксплуатационным требованиям несмотря на то, что они технически исправны, что обычно наблюдается в устаревании систем контроля, когда их невозможно интегрировать с новыми технологиями; выявляются проблемы с соблюдением законодательных и договорных требований; отсутствует оценка совместимости новых запасных частей и старого оборудования.

Риски экономического устаревания становятся очевидными, когда затраты и усилия, необходимые для поддержания актива в рабочем состоянии, превышают ценность, которую он приносит, что часто вынуждает организации принимать реактивные решения о капиталовложениях; возникают потери, связанные с простоем; в организации отсутствует поставщик.

В реальности эти категории риска редко существуют изолированно, и эффективное управление зависит от осознания их взаимодействия, а не от рассмотрения их как отдельных условий.

Почему управление устареванием не удается в некоторых организациях

В различных отраслях промышленности неудачи в управлении устареванием могут быть связаны с последовательными и наблюдаемыми закономерностями, а не с отдельными проблемами. Наиболее распространенный сбой начинается с глубоко укоренившегося реактивного мышления, при котором активы эксплуатируются до отказа, и в этот момент организация вынуждена реагировать с ограниченными опциями, более высокими затратами и повышенным операционным риском. Такое поведение часто усугубляется неполными или некорректными данными об активах в компьютеризованных системах, где неполные спецификации, отсутствующая информация о производителе и некорректное

управление данными препятствуют раннему выявлению рисков устаревания.

Еще одна проблема возникает из-за недостаточного понимания персоналом жизненного цикла объекта, когда организация систематически не отслеживает статус работы поставщиков или этапы жизненного цикла, что приводит к запоздалому информированию о прекращении работы или отзыве технической поддержки. Функции планирования и составление графиков часто не связаны с управлением устареванием, в результате чего соответствующая работа рассматривается как срочная, а не запланированная и оптимизированная. Функции закупок, как правило, включаются в процесс слишком поздно, что ограничивает возможности поиска поставщиков и увеличивает зависимость от ускоренных или нестандартных решений.

Возможно, наиболее критичным является то, что зачастую отсутствует четкая ответственность в организации за риск устаревания оборудования, а обязанности распределены между инженерными подразделениями, службами технического обслуживания и подразделениями поставок и логистики, что приводит к задержкам, несогласованности и быстрым, но неоптимальным решениям. Системы ключевых показателей эффективно, там, где они существуют, как правило, фокусируются на реактивных показателях, таких как затраты или объем невыполненной работы, а не на проактивной оценке подверженности риску и готовности к нему. Совокупный эффект этих факторов заключается в том, что система постоянно реагирует на давление, а не действует на опережение.

Эффективное управление устареванием требует структурированного подхода, в соответствии с циклической моделью принятия решений Уильяма Эдвардса Деминга, используемая для управления процессами и системами. **«Планирование-Действие-Проверка-Корректировка» — это цикл Шухарта-Деминга**, гарантирующий, что идентификация, принятие решений, их исполнение и обратная связь интегрированы в обычные операционные процессы, а не рассматриваются как отдельные мероприятия.

Далее мы приведем несколько примеров мероприятий (включая, но не ограничиваясь) на каждый из циклов модели Уильяма Эдвардса Деминга.

Этап «Планирование» (Plan)

Целью данного этапа является идентификация и оценка рисков устаревания, разработка стратегии управления для минимизации негативных последствий в рамках жизненного цикла оборудования.

Типовые действия могут включать:

1. Формирование межфункциональной рабочей группы (инжиниринг, техническое обслуживание, руководство) для координации усилий по управлению устареванием.
2. Проведение исторического анализа данных по планово-предупредительному ремонту для выявления закономерностей и трендов устаревания.
3. Идентификацию критически важных компонентов оборудования и определение их жизненного цикла (включая даты окончания производства, поддержки и поставок).
4. Разработку комплексной стратегии цепочки поставок, включающей применение подхода «каннибализации» оборудования (использование компонентов с выведенных из эксплуатации единиц).
5. Создание и актуализацию баз данных для отслеживания состояния оборудования, сроков устаревания и доступности компонентов.
6. Согласование бюджета на мероприятия по управлению устареванием (закупки, модернизация, обучение персонала и т.д.).
7. Планирование регулярных проверок состояния оборудования.
8. Мониторинг рынка и поставщиков: отслеживание изменений в продуктовых линейках, прекращении производства и поддержке компонентов.

Мы полагаем, что ожидаемый управленческий эффект этапа «Планирование» включает в себя снижение неопределённости в отношении рисков устаревания, формирование базы данных для принятия обоснованных решений, оптимизацию и повышение предсказуемости затрат и сроков выполнения работ на управление устареванием за счёт превентивных мер и минимизацию вероятности внезапных простоев из-за отсутствия запчастей.

Этап «Действие» (Do)

Целью данного этапа является: реализация разработанных стратегий и мероприятий по снижению рисков устаревания в соответствии с утверждённым планом.

Типовые действия могут включать в себя:

1. Детальный анализ рисков устаревания с участием межфункциональной группы, включая количественную оценку вероятности и последствий.
2. Внедрение системы автоматических оповещений о компонентах с истекающим сроком службы или поддержки.
3. Обеспечение наличия альтернативных поставщиков и резервных источников поставок.
4. Разработку и выполнение планов тестирования

критически важных систем (Проверка совместимости новых компонентов со старым оборудованием, оценка влияния обновлений на стабильность работы).

5. Проведение долгосрочных тестов производительности и надёжности оборудования в штатных и экстремальных условиях.
6. Создание централизованной базы знаний по управлению устареванием:
 - Документирование лучших практик;
 - Фиксация извлечённых уроков;
 - Сбор данных о случаях устаревания и способах их устранения.
7. Разработку и внедрение коммуникационного плана для обеспечения взаимодействия между подразделениями.
8. Привлечение технических экспертов для принятия сложных решений (модернизация, замена, адаптация).

Ожидаемый управленческий эффект этапа «Действие» состоит в снижении вероятности незапланированных простоев, в сокращении времени реагирования на происшествя в улучшении взаимодействия между подразделениями и накоплении знаний для дальнейшего совершенствования процессов.

Этап «Проверка» (Check)

Целью данного этапа является мониторинг эффективности реализованных мероприятий, выявление отклонений и анализ причин их возникновения.

Типовые действия могут включать в себя:

1. Оценку результативности стратегий смягчения последствий (закупка перед прекращением производства, перепроектирование, альтернативные источники).
2. Проведение регулярных совещаний (например, ежеквартально) с рабочей группой для анализа текущей ситуации и корректировки планов.
3. Обновление оценки рисков с учётом изменений на рынке и у поставщиков.
4. Аудит эффективности работы поставщиков.
5. Сравнение уровня запасов с прогнозируемым спросом и корректировка стратегии закупок.
6. Проверку условий хранения запасных частей и компонентов.
7. Анализ тенденций потребления критически важных компонентов для прогнозирования будущих потребностей.
8. Подготовку структурированных отчётов об устаревании оборудования:
9. Использование информационных панелей и визуальной аналитики для наглядного представления данных.

10. Оценку затрат, связанных с перепроектированием, срочными закупками и задержками.
11. Мониторинг проактивных и реактивных КПЭ, связанных с устареванием.
12. Анализ прошлых случаев устаревания для выявления системных проблем.
13. Проверку соответствия процессов внутренним и отраслевым стандартам.

Ожидаемый управленческий эффект этапа «Проверка» состоит в повышении точности прогнозирования рисков устаревания и оптимизации запасов и затрат на закупки, в своевременном выявлении и устранении проблем в цепочке поставок, в улучшении качества данных для принятия решений и снижении доли реактивных решений за счёт раннего обнаружения проблем.

Этап «Корректировка» (Act)

Целью данного этапа является внесение изменений в процессы и стратегии на основе результатов мониторинга для постоянного совершенствования системы управления устареванием.

Типовые действия могут включать в себя:

1. Стандартизацию успешных изменений в дизайне оборудования, процессах и управлении запасами.
2. Пересмотр стратегий устранения устаревания с учётом реальных сбоев и извлечённых уроков.
3. Обновление матриц рисков и допущений при планировании будущих проектов.
4. Корректировку спецификаций материалов с учётом их жизненного цикла и доступности.
5. Перепроектирование компонентов с высокой степенью риска устаревания (замена на более долговечные или совместимые аналоги).
6. Устранение повторяющихся факторов, приводящих к устареванию оборудования (например, закупка компонентов коротким жизненным циклом).
7. Обеспечение стратегического охвата запасными частями:
 - Формирование долгосрочных запасов;
 - Заключение долгосрочных контрактов с поставщиками.
8. Включение обязанностей по управлению устареванием в организационную структуру (назначение ответственных, распределение ролей).
9. Обучение персонала новым процессам и стандартам работы.
10. Интеграцию улучшений в систему KPI и отчетность.

Ожидаемый управленческий эффект этапа «Корректировка» состоит в повышении предсказуемости затрат и сроков эксплуатации оборудования, в улучшении экономических и операционных показателей и создании

устойчивой системы, способной адаптироваться к изменениям рынка и технологий.

Хорошими примерами «эталона успеха» на этапе «Корректировка» могут быть следующие результаты:

- Для более 90 % компонентов, подверженных высокому риску, разработаны планы по снижению рисков
- Время простоя, связанное с устареванием оборудования, сокращается из года в год
- Расходы на экстренные закупки составляют менее 5 % от расходов на техническое обслуживание
- Наличие критически важных запасных частей превышает 95 %
- Все проекты последовательно следуют обновленным процедурам

Практическая основа для разработки ключевых показателей эффективности для управления устареванием

Разработка эффективных ключевых показателей эффективности для управления устареванием начинается с четкого определения бизнес-целей, которые обычно включают повышение надежности, снижение операционных рисков и достижение предсказуемости затрат. Без такой основы ключевые показатели эффективности рискуют превратиться в «показатели тщеславия», а не в стимулы к действию.

Источники данных должны быть четко определены и подтверждены, как правило, это компьютеризированная система управления техническим обслуживанием для получения данных об оборудовании и заказах на выполнение работ, системы инвентаризации для определения уровня запасов и сроков выполнения заказов, базы данных поставщиков для получения информации о жизненном цикле и учет технических спецификаций. Признавая, что качество данных изначально будет недостаточным, организациям следует сосредоточиться на установлении исходных условий и повышении точности с течением времени, а не откладывать внедрение.

Установление пороговых значений и целевых показателей требует баланса между амбициями и реализмом, а также понимания текущего уровня зрелости системы технического обслуживания и операционных ограничений. Целевые показатели должны отражать то, что достижимо в контексте конкретной организации с конкретными активами.

В таблице ниже приведены примеры КПЭ (табл. 1), которые можно использовать при разработке сбалансированной системы показателей, он содержит краткие определения и формулы для мониторинга и отслеживания эффективности управления устареванием.

Таблица 1.

Ключевые показатели эффективности

#	КПЭ	Категории	Тип	Причина типизации	Определение	Формула	Единица	Источник данных
1	Процент контролируемых компонентов	Оперативный	Проактивный	1. Отражает способность организации заранее выявлять риски устаревания за счёт систематического мониторинга компонентов. 2. Позволяет preventивно планировать мероприятия по управлению устареванием (закупки, создание запасов, поиск аналогов). 3. Направлен на предотвращение аварийных ситуаций и незапланированных простоев, а не на их ликвидацию. 4. Связан с прогнозированием будущих проблем, а не с реакцией на уже произошедшие события. 5. Влияет на стратегическое планирование жизненного цикла активов, а не на оперативное устранение последствий	Доля критических компонентов, активно отслеживаемых на предмет устаревания.	$(\text{Количество отслеживаемых критических компонентов} / \text{общее количество критических компонентов}) \times 100$	%	1. Компьютеризованная система управления ТОиР (CMMS / EAM-система). 2. Реестр критически важных компонентов с указанием: — сроков окончания производства; — дат прекращения технической поддержки; — доступности запасных частей; — альтернативных поставщиков; — планов замены/модернизации
2	Частота обновления базы данных	Оперативный	Проактивный	1. Отражает способность организации preventивно поддерживать актуальность данных о рисках устаревания. 2. Позволяет заранее выявлять пробелы в информации и оперативно их устранять до возникновения критических ситуаций. 3. Направлен на предотвращение последствий использования устаревших данных (ошибочные решения, задержки в закупках, незапланированные простои). 4. Связан с систематическим поддержанием информационной базы в актуальном состоянии. 5. Влияет на качество прогнозирования и планирования мероприятий по управлению устареванием	Как часто в системе обновляются записи об устаревании.	$\text{Количество обновлений базы данных по устареванию, выполненных за определенный период (необязательный процент соответствия} = \frac{\text{Фактические обновления / запланированные обновления}}{\text{x 100}}$	Количество / %	1. Логги системы управления техническим обслуживанием и ремонтами (CMMS / EAM-система) — журнал изменений и обновлений. 2. Регламент управления данными по устареванию оборудования — план-график актуализации информации. 3. Отчёты ответственных лиц о выполнении плановых обновлений базы данных
3	Точность прогноза	Техническая	Проактивный	1. Отражает способность организации заранее выявлять и прогнозировать риски устаревания до их реализации. 2. Позволяет preventивно планировать мероприятия по управлению устареванием (закупки стратегических запасов, поиск аналогов, модернизацию). 3. Направлен на предотвращение незапланированных простоев, срочных закупок и аварийных ремонтов. 4. Связан с прогнозированием будущих проблем на основе анализа жизненного цикла компонентов, рыночных тенденций и данных поставщиков. 5. Влияет на стратегическое планирование бюджета и ресурсов, снижая долю реактивных и дорогостоящих решений	Точность прогнозов устаревания по сравнению с фактическими случаями устаревания.	$(\text{Количество правильно спрогнозированных случаев устаревания} / \text{общее количество прогнозируемых случаев}) \times 100$	%	1. Архив прогнозов устаревания — база данных с зафиксированными прогнозными (компонент, дата прогноза, ожидаемая дата устаревания, причины). 2. Журнал зафиксированных случаев устаревания — данные о фактически наступивших случаях (компонент, дата наступления, причины, последствия). 3. CMMS / EAM-система — записи о ремонтах, заменах и сбоях, вызванных устареванием. 4. Отчёты подразделений ТО и снабжения — информация о проблемах с поставками, поддержкой и совместимостью компонентов

Продолжение таблицы

#	КПЭ	Категории	Тип	Причина типизации	Определение	Формула	Единица	Источник данных
4	Уровень резервных запасов	Экономическая	Проактивный	- Позволяет заранее оценить готовность организации к рискам устаревания до возникновения дефицита компонентов. - Даёт возможность превентивно корректировать планы закупок и формировать стратегические запасы. - Направлен на предотвращение незапланированных простоев из-за отсутствия запчастей и срочных закупок по завышенным ценам. - Связан с прогнозированием потребностей на основе жизненного цикла компонентов, анализа поставщиков и рыночных тенденций. - Влияет на стратегическое планирование бюджета и логистических процессов, снижая долю реактивных и дорогостоящих решений	Доступность стратегических резервных запасов для под- верженных риску или устаревших компонентов.	Текущий запас критически важных резервных компонентов (шт) / Целевой уровень запаса (шт) x 100%	% (процент от целевого уровня запаса)	- Система управления запасами и складом — актуальные данные о количестве компонентов на складе. - Реестр критически важных компонентов с указанием целевого уровня запасов (утверждённый норматив). - СМMS / EAM-система — данные о критичности компонентов и прогнозных сроках устаревания. - Планы закупок и стратегии управления устареванием — обоснование целевых уровней запасов для каждого компонента. - Отчёты службы снабжения — информация о доступности компонентов у поставщиков и сроках поставок
5	Время на замену компонента	Оперативный	Реактивный	- Фиксирует уже произошедшие события — время, затраченное на устранение последствий устаревания. - Измеряет реакцию организации на возникшую проблему (износ/устаревание) - Отражает последствия недостаточного проактивного управления (отсутствие запасов, неготовность к замене, задержки в закупках). - Используется для анализа и устранения причин задержек после их возникновения, а не для прогнозирования рисков. - Связан с ликвидацией последствий устаревания, а не с его предотвращением через планирование и мониторинг	Среднее время, затрачиваемое на полную замену компонента от момента фиксации его морального или физического износа до успешного ввода в эксплуатацию нового/модернизированного элемента	Сумма (дата завершения замены — дата определения морального износа) / Количество выполненных замен	Дни	- СМMS / EAM-система — записи о заявках на замену, датах фиксации износа и датах завершения работ. - Журналы учёта технического обслуживания и ремонтов — данные о времени выполнения операций. - Системы управления запасами — информация о времени ожидания поставки компонентов. - Табели учёта рабочего времени — данные о затраченных человеко-часах. - Отчёты исполнителей и руководителей ремонтных бригад — детализация этапов замены и задержек

6	Показатель успешности внедрения замены	Оперативный	Реактивный	1. Фиксирует уже произошедшие события — результаты выполненных замен компонентов. 2. Измеряет реакцию системы на проведённые изменения, а не способность предотвратить проблемы заранее. 3. Отражает последствия недостаточно проработанных решений по замене (несовместимость, ошибки монтажа, некачественные компоненты). 4. Используется для анализа и устранения причин неудач после их возникновения, а не для прогнозирования рисков. 5. Связан с ликвидацией последствий неудачных замен (доработки, повторные замены), а не с их предотвращением через планирование и тестирование	Доля успешно выполненных замен компонентов, при которых не потребовалось дополнительных доработок, не возникло функциональных сбоев и не нарушилась работа сопряжённых систем	(Количество успешных замен / Общее количество осуществлённых замен) x 100	%	1. CMMS / EAM-система — записи о выполненных заменах, статусах и результатах внедрения. 2. Журналы ТО и ремонтов — фиксация случаев сбоев после замены. 3. Отчёты испытательных подразделений — результаты тестирования оборудования после замены компонентов. 4. Система управления качеством — данные о выявленных дефектах и доработках. 5. Обращения пользователей/операторов оборудования — информация о нарушениях функциональности после замены. 6. Отчёты руководителей ремонтных бригад — детализация причин неудачных замен
7	Соответствие техническим требованиям	Технический	Реактивный	1. Фиксирует уже произошедшие события — результаты проверок после выполнения замен. 2. Измеряет реакцию системы на проведённые изменения, а не способность предотвратить несоответствия заранее. 3. Отражает последствия недостаточно строгого входного контроля, ошибок в закупках или отклонений от утверждённых процедур монтажа. 4. Используется для анализа и устранения причин несоответствий после их выявления, а не для прогнозирования рисков. 5. Связан с ликвидацией последствий установки несоответствующих компонентов (перемонтаж, доработки, внеплановые проверки), а не с их предотвращением через планирование и контроль качества	Доля замен компонентов, полностью соответствующих утверждённому техническим спецификациям, стандартам организации и отраслевым нормативам (включая параметры совместимости, надёжности, безопасности и эксплуатационные характеристики)	(Количество замен, соответствующих всем техническим требованиям / Оценено общее количество замен) x 100	%	1. Протоколы испытаний и приёмы оборудования — результаты проверки соответствия после замены. 2. CMMS / EAM-система — записи о выполненных заменах и их статусе соответствия требованиям. 3. Техническая документация и спецификации — утверждённые требования к компонентам (чертежи, ТУ, ГОСТы, ISO и т.д.). 4. Отчёты службы технического контроля и ОТК — данные о выявленных отклонениях от требований. 5. Журналы входного контроля — результаты проверки комплектующих перед установкой. 6. Система управления качеством — регистрация несоответствий и корректирующих действий. 7. Акты аудитов и инспекций — заключение внутренних и внешних проверок соответствия стандартам

Продолжение таблицы

#	КПЭ	Категории	Тип	Причина типизации	Определение	Формула	Единица	Источник данных
8	Затраты на управление устареванием	Экономическая	Реактивный	1. Фиксирует уже понесённые расходы, связанные с управлением устареванием. 2. Измеряет реакцию организации на возникшие проблемы (срочные закупки, экстренные ремонты, привлечение внешних экспертов), а не способность их предотвратить. 3. Отражает последствия недостаточного проактивного управления (высокие затраты на экстренные меры). 4. Используется для анализа и оптимизации бюджета на следующий период, а не для прогнозирования рисков. 5. Связан с ликвидацией последствий устаревания (ремонт, замена, срочные закупки), а не с их предотвращением через планирование и мониторинг	Совокупные фактические расходы организации за отчётный период, связанные с реализацией мероприятий по управлению устареванием оборудования. В состав затрат включаются: — оплата труда персонала, задействованного в процессах управления устареванием (включая обучение и повышение квалификации); — затраты на приобретение, внедрение и сопровождение специализированного ПО и инструментов мониторинга; — расходы на администрирование процессов (планирование, отчётность, координация); — услуги внешних консультантов и подрядчиков по вопросам управления устареванием; — создание и поддержание стратегических запасов критически важных компонентов	$ЗУ = 3(\text{перс}) + 3(\text{по}) + 3(\text{адм}) + 3(\text{внеш}) + 3(\text{зап})$ ЗУ — затраты на управление устареванием, руб.; З(перс) — затраты на персонал, руб. (ФОТ + обучение); З(по) — затраты на ПО и инструменты, руб.; З(адм) — затраты на администрирование, руб.; З(внеш) — затраты на внешнюю поддержку, руб.; З(зап) — затраты на формирование и хранение запасов, руб.	В валюте (руб)	1. Бухгалтерская отчётность — данные по статьям затрат. 2. Бюджеты подразделений — плановые и фактические расходы на управление устареванием. 3. CMMS / EAM-система — затраты на ТОиР, связанные с устареванием. 4. Система управления запасами — расходы на формирование стратегических запасов. 5. Договоры с внешними подрядчиками и консультантами — стоимость услуг по управлению устареванием. 6. Отчёты отдела кадров — затраты на обучение и зарплаты персонала. 7. Договоры на ПО и лицензии — расходы на программное обеспечение. 8. Внутренние регламенты и планы — бюджеты на администрирование и координацию процессов

Заключение

Целью исследования являлась разработка риск-ориентированной концептуальной модели управления устареванием оборудования в ресурсоёмких отраслях, объединяющую классификацию видов устаревания, анализ причин организационной неэффективности, цикл PDCA и систему ключевых показателей эффективности для мониторинга и снижения риска.

Научная новизна проведённого исследования заключается в следующем:

1. Обосновано понимание устаревания оборудования как системного управленческого риска (а не только технического следствия износа).
2. Предложена трёхкомпонентная классификация устаревания оборудования, включающая технический, функциональный и экономический аспекты, рассматриваемые во взаимосвязи.
3. Систематизированы организационные причины неэффективного управления устареванием, включая реактивный стиль управления, недостаточную полноту данных, размытость ответственности и преобладание реактивных показателей.
4. Разработана концептуальная риск-ориентированная модель управления устареванием оборудования на основе цикла PDCA, интегрирующая идентификацию риска, принятие решений, реализацию мер, мониторинг и корректирующие действия.
5. Предложена система КПЭ для оценки состояния и результативности управления устареванием, включающая как проактивные, так и реактивные показатели, что позволяет комплексно оценивать состояние и результативность управления устареванием.

Основные результаты исследования:

1. Подтверждена гипотеза о том, что эффективность управления устареванием существенно возрастает при рассмотрении его как системного риска и применении интегрированного подхода.
 - Показано, что внедрение модели PDCA в процессы управления позволяет снизить долю реактивных решений, повысить предсказуемость отказов и затрат и улучшить межфункциональную координацию (инжиниринг, ТОиР, снабжение и пр).
2. Показана практическая целесообразность применения предложенной системы КПЭ для:
 - Раннего выявления компонентов, подверженных риску устаревания;
 - Планирования мероприятий по замене/модернизации;
 - Снижения доли срочных решений;
 - Повышения устойчивости производственных процессов.
3. Определены ключевые условия успешной реализации модели:
 - Формирование межфункциональных рабочих групп;
 - Актуализация баз данных по жизненному циклу активов и поставщиков;
 - Внедрение автоматических оповещений о рисках устаревания;
 - Стандартизация процедур мониторинга и корректирующих действий.

Таким образом, разработанная модель создаёт методическую основу для перехода от фрагментарного реагирования к системному риск-ориентированному управлению устареванием. Её внедрение позволит предприятиям повысить надёжность эксплуатации оборудования, снизить операционные потери и повысить обоснованность управленческих решений в условиях технологической и рыночной неопределённости.

ЛИТЕРАТУРА

1. Romero Rojo F.J., Roy R., Shehab E. Obsolescence management for long-life contracts: state of the art and future trends // The International Journal of Advanced Manufacturing Technology. 2010. Vol. 49. P. 1235–1250. DOI: 10.1007/s00170-009-2471-3.
2. Liu Y., Zolghadri M. What is happening in product obsolescence management? Three decades of review from 1993 to 2023 // Research in Engineering Design. 2026. Vol. 37. Art. 9. DOI: 10.1007/s00163-026-00470-8.
3. Ferreira N.N., Martins M.R., Dziedzic R., Figueiredo M.A.G. de. Obsolescence management for offshore oil and gas production facility life extension // Ocean Engineering. 2023. Vol. 285. Art. 115388. DOI: 10.1016/j.oceaneng.2023.115388.
4. Ghaithan A.M., AlShamrani O., Mohammed A., Alshibani A. Proactive framework for obsolescence management of electrical equipment in oil and gas industry // Journal of Quality in Maintenance Engineering. 2024. Vol. 30, no. 3. P. 493–507. DOI: 10.1108/JQME-10-2023-0096.
5. Ault T., Bradley T.H. Risk-based approach for managing obsolescence for automation systems in heavy industries // Systems Engineering. 2022. Vol. 25, no. 6. P. 551–560. DOI: 10.1002/sys.21635.
6. Ferreira S., Silva F.J.G., Casais R.B., Pereira M.T., Ferreira L.P. KPI development and obsolescence management in industrial maintenance // Procedia Manufacturing. 2019. Vol. 38. P. 1427–1435. DOI: 10.1016/j.promfg.2020.01.145.
7. Souifi A., Zolghadri M. Obsolescence Risk assesment: Key Challenges and Practical Lessons // IFAC-PapersOnLine. 2024. Vol. 58, no. 19. P. 486–491. DOI: 10.1016/j.ifacol.2024.09.259.
8. Liu Y., Karaani S., Zolghadri M. An Approach for Implementing Obsolescence Management Strategies to Extend Product Service Lifespan // Procedia CIRP. 2025. Vol. 135. P. 660–665. DOI: 10.1016/j.procir.2025.01.067.

9. Collier Z.A., Lambert J.H. Managing obsolescence of embedded hardware and software in secure and trusted systems // *Frontiers of Engineering Management*. 2020. Vol. 7, no. 2. P. 172–181. DOI: 10.1007/s42524-019-0032-5.
10. Konoza B., Sandborn P. Evaluating the End of Maintenance Dates for Electronic Assemblies Composed of Obsolete Parts // *Journal of Mechanical Design*. 2014. Vol. 136. DOI: 10.1115/1.4026096.
11. ГОСТ Р МЭК 62402-2023 — национальный стандарт Российской Федерации «Надёжность в технике. Управление устареванием». ГОСТ Р 70383-2022 — национальный стандарт Российской Федерации «Системы промышленной автоматизации и интеграция. Управление устареванием. Рекомендации по управлению устареванием для пользователей и операторов на промышленных предприятиях».
12. ГОСТ Р 27.203–2012 «Надёжность в технике. Управление устареванием». Федеральные нормы и правила в области использования атомной энергии (НП-096–15).
13. IAEA Safety Rep. Ser. № SRS-109 «Regulatory Oversight of Ageing Management and Long Term Operation Programme of Nuclear Power Plants» (МАГАТЭ, 2022).

© Клементьева Ольга Васильевна; Чмилевская Оксана Сергеевна
Журнал «Современная наука: актуальные проблемы теории и практики»

КОЛИЧЕСТВЕННАЯ ОЦЕНКА ВЛИЯНИЯ ПОНИЖЕНИЯ РАЗРЕШЕНИЯ РАСТРОВЫХ ИЗОБРАЖЕНИЙ НА ЭТАПЕ ДЕКОДИРОВАНИЯ НА ПОТРЕБЛЕНИЕ ПАМЯТИ, ПОВЕДЕНИЕ КЕША И ПЛАВНОСТЬ ИНТЕРФЕЙСА ВО FLUTTER

QUANTATIVE ASSESSMENT
OF THE IMPACT OF DECODE-TIME
DOWNSCALING OF RASTER IMAGES
ON MEMORY CONSUMPTION, IMAGE
CACHE BEHAVIOUR AND USER INTERFACE
SMOOTHNESS IN FLUTTER

A. Kozlov
M. Kochemirov
D. Ushakov

Summary. The object of research is raster images in mobile applications on the Flutter platform, and the subject is the reduction of their resolution at the decoding stage via the `cacheWidth/cacheHeight` parameters and the `ResizeImage` provider. The method combines microbenchmarks, queries to the image cache API and per-frame `FrameTiming` measurements on a physical device in profile mode, with the measurements duplicated on two renderers — Impeller (Vulkan) and Skia (OpenGL ES). The decoded buffer size is shown to be strictly quadratic in the downscaling factor, reaching a 1600-fold saving, whereas the PNG decoding time is practically independent of this factor: the benefit of decode-time resizing lies in memory and cache capacity rather than in CPU load. On cache overflow, native decoding causes 20 evictions and re-decoding of every revisited element versus zero under resizing, while the number of raster frames exceeding the budget is reduced roughly fourfold on both renderers.

Keywords: Flutter, image decoding, `cacheWidth`, image cache, memory consumption, UI smoothness, dropped frames, rasterization, Impeller, mobile development.

Козлов Антон Вячеславович

Санкт-Петербургский государственный университет
аэрокосмического приборостроения
antgress@gmail.com

Кочемиров Максим Алексеевич

Санкт-Петербургский государственный университет
аэрокосмического приборостроения
kma-2004@mail.ru

Ушаков Даниил Павлович

Санкт-Петербургский государственный университет
аэрокосмического приборостроения
ushdaniil@yandex.ru

Аннотация. Объектом исследования являются растровые изображения в мобильных приложениях на платформе Flutter, предметом — понижение их разрешения на этапе декодирования посредством параметров `cacheWidth/cacheHeight` и провайдера `ResizeImage`. Методика сочетает микротесты производительности, опрос программного интерфейса кэша изображений и измерение кадровых временных характеристик `FrameTiming` на физическом устройстве в режиме `profile`, причём измерения продублированы на двух рендерерах — Impeller (Vulkan) и Skia (OpenGL ES). Установлено, что объём декодированного буфера строго квадратичен по коэффициенту понижения разрешения и достигает экономии в 1600 раз, тогда как время декодирования формата PNG от этого коэффициента практически не зависит: выигрыш понижения разрешения на этапе декодирования сосредоточен в памяти и ёмкости кэша, а не в нагрузке на процессор. При переполнении кэша декодирование в полном разрешении вызывает 20 вытеснений и повторное декодирование всех возвращаемых элементов против нуля при понижении разрешения, а число растровых кадров сверх бюджета сокращается примерно вчетверо на обоих рендерерах. На основании полученных данных сформулированы практические рекомендации по выбору уровня и способа понижения разрешения.

Ключевые слова: Flutter, декодирование изображений, `cacheWidth`, кэш изображений, потребление памяти, плавность интерфейса, пропуск кадров, растеризация, Impeller, мобильная разработка.

Введение

Мобильные приложения на Flutter повседневно отображают коллекции из десятков и сотен растровых изображений. По умолчанию среда декодирует изображение в полном исходном разрешении в буфер RGBA8888 объёмом $W \cdot H \cdot 4$ байта, кэширует его и лишь затем масштабирует силами графическо-

го процессора под размер виджета [3, 5]; фотография 4000×3000 занимает при этом около 48 МБ независимо от размера показа. Параметры `cacheWidth/cacheHeight` и провайдер `ResizeImage` переносят понижение разрешения на этап декодирования, помещая в кэш уже уменьшенный буфер [1, 2]; отладочный инструментарий Flutter штатно помечает изображения с избыточным разрешением и рекомендует эти средства [1]. Механизм

документирован, но качественно: измеренной зависимости времени декодирования от коэффициента понижения, связи переполнения кэша с повторным декодированием и влияния понижения разрешения на пропуск кадров в литературе, особенно русскоязычной, нет.

Цель работы — количественно оценить влияние понижения разрешения на этапе декодирования на память, время декодирования, переполнение кэша и плавность прокрутки на реальном Android-устройстве, причём все измерения продублированы на обоих рендерерах Flutter — Impeller (Vulkan) и Skia (OpenGL ES), что отделяет эффекты работы с изображениями от артефактов графической подсистемы. Задачи: разработать сценарии, изолирующие отдельные эффекты; провести измерения с контролем помех и статистической обработкой; объяснить наблюдаемое через конвейер декодирования, кэширования и растеризации.

Теоретическое обоснование

1. Конвейер декодирования и хранения изображений. Байты ассета поступают в `instantiatelImageCodec` и декодируются в несжатый буфер `RGBA8888` [3], который кэшируется, загружается на графический процессор текстурой и участвует в композиции кадра [5]. Объём такого буфера не зависит от сжатия исходного файла и равен

$$S = W \cdot H \cdot 4 \text{ байт}, \quad (1)$$

где W и H — ширина и высота в пикселях, 4 — число байт на пиксель в `RGBA8888`. Декодирование идёт вне UI-потока, а загрузка текстуры и композиция — в потоке растеризации [5].

2. Кэш изображений. Кэш Flutter работает по схеме вытеснения наименее недавно использованного элемента (LRU) и по умолчанию хранит до 1000 записей объёмом до 100 МиБ, то есть 104 857 600 байт [4]. Ключ записи включает `cacheWidth/cacheHeight` и масштаб, поэтому один ассет в разных целевых размерах занимает разные записи; вытесненное изображение при повторном показе декодируется заново [4]. Значимость кэширования для производительности отмечается и в отечественной литературе [13].
3. Понижение разрешения на этапе декодирования. Значения `cacheWidth/cacheHeight` передаются в `instantiatelImageCodec` как `targetWidth/targetHeight`, а `ResizeImage` — обёртка над тем же механизмом [1, 2, 3]; буфер получает размер $W_{ц} \cdot H_{ц} \cdot 4$, квадратичный по линейному коэффициенту. Уменьшает ли кодек изображение в ходе декодирования или декодирует целиком с после-

дующей передискретизацией, зависит от формата: JPEG допускает декодирование с уменьшением в долях 1/2, 1/4, 1/8 на уровне обратного дискретного косинусного преобразования [10], тогда как PNG режима масштабированного декодирования не имеет. Это различие оказывается ключевым для результатов по времени декодирования.

4. Бюджет кадра, фазы построения и растеризации. При частоте 60 Гц бюджет кадра равен 16,67 мс, при 120 Гц — 8,33 мс [6, 7]. Кадр состоит из фазы построения (код Dart в UI-потоке, формирование дерева слоёв) и фазы растеризации (загрузка текстур и композиция на графическом процессоре) [5, 7]; превышение бюджета любой фазой даёт потерянный кадр, воспринимаемый как рывок [7]. Класс `FrameTiming` отдельно сообщает `buildDuration` и `rasterDuration` [6], а методика измерения производительности Flutter рекомендует отслеживать среднее, 90-й и 99-й перцентили и худший кадр [8].
5. Рендереры Impeller и Skia. С версии Flutter 3.27 рендерером по умолчанию является Impeller на Vulkan (Android API 29 и выше) с откатом к устаревшему Skia на OpenGL ES; флаг `--no-enable-impeller` принудительно выбирает Skia [9]. Декодирование в обоих случаях выполняет один и тот же кодек, поэтому от рендерера зависит не оно, а путь растеризации — что и обуславливает изменения на обоих рендерерах.

Методика эксперимента

1. Сценарии тестирования. Пять сценариев изолируют по одному эффекту: I1 — объём декодированного буфера в зависимости от целевого разрешения (проверка); I2 — время декодирования в зависимости от коэффициента понижения; I3 — вытеснения и повторное декодирование при галерее, превышающей бюджет кэша; I4 — времена растеризации при программной прокрутке сетки; I5 — сравнение способов понижения разрешения.
2. Конфигурация. Смартфон HONOR FNE-NX9 (физический), Android 14 (API 34), экран 1080×2400, плотность пикселей 3,0, частота 60 Гц (бюджет кадра 16,67 мс); Flutter 3.38.4, Dart 3.10.3, режим `profile` (`debug`, симуляторы и эмуляторы исключены — иначе измерения недостоверны [5]). Сняты два набора: основной на Impeller (Vulkan, по умолчанию) и контрольный на Skia (`--no-enable-impeller`) [9]; фактические лимиты кэша — 104 857 600 байт и 1000 записей [4]. Изображения локальные, сгенерированы программно в формате PNG (единственный встроенный во Flutter кодировщик для генерации ассетов); сеть исключена.
3. Метрики и инструменты. Память — `imageCache.currentSizeBytes` с проверкой против теорети-

ческого W-H-4 [4]. Время декодирования — секундомер вокруг пары `instantiatedImageCodec` → `getNextFrame` при заранее загруженных байтах, без файлового ввода-вывода и компоновки [3]. Поведение кэша — `currentSizeBytes`, `currentSize`, `liveImageCount`, число вытеснений (разность общего числа декодирований и текущего числа записей) и повторных декодирований по инструментированному счётчику [4]. Кадры — подписка `addTimingsCallback` с разбором `buildDuration` и `rasterDuration`, бюджет из фактической частоты, подсчёт кадров сверх бюджета и перцентилей 50/90/99 и максимума [6, 8].

4. Статистическая обработка и контроль помех. Для каждой точки выполнялось пять измеряемых прогонов и один прогрев; минимум и максимум отбрасывались, бралась медиана оставшихся [11], перцентили — методом ближайшего ранга. Между прогонами кэш очищался (`clear`, `clearLiveImages`) [4], выдерживались паузы не менее 500 мс (между сценариями — не менее двух секунд) против теплового троттлинга. Фактическое разрешение декодирования проверялось на каждой загрузке (нарушений за оба набора нет); тело цикла защищено от удаления оптимизатором направлением результатов в приёмник [12].
5. Отклонения от плана и оговорки. Использован формат PNG, а не JPEG (встроенного JPEG-кодировщика для генерации нет): вывод о стоимости декодирования справедлив для PNG, для JPEG ожидается иное в силу масштабированного декодирования [10]. Сценарий I4 выполнен на исходниках 2000×1500, а не 4000×3000: при трёх колонках в видимой области удерживается порядка 20–27 живых изображений, что при 4000×3000 дало бы около 1,3 ГБ и нехватку памяти; I3 — на полных 4000×3000. Скорость прокрутки 533 пикс./с отвечает быстрому пролистыванию ленты (при 333 пикс./с эффект был на грани шума) и одинакова в обоих вариантах. Измерения проведены на одном устройстве с частотой 60 Гц.

Результаты и обсуждение

1. Сценарий I1: память как функция коэффициента понижения разрешения (проверка).

Во всех 15 точках (три исходника × пять целей) фактический объём в кэше совпал с теоретическим W-H-4 до байта при нулевом разбросе между прогонами, а декодированное разрешение всегда равнялось целевому — понижение происходит именно на этапе декодирования. Зависимость строго квадратична: для исходника 4000×3000 буфер полного разрешения занимает 48 000 000 байт, а цели 800/400/200/100 пикс. дают соответственно 1 920 000, 480 000, 120 000 и 30 000 байт,

то есть при уменьшении стороны в 40 раз память сокращается в 1600 раз. Сценарий служит опорной проверкой и подтверждает корректность измерительной методики.

2. Сценарий I2: для PNG понижение разрешения не экономит процессорное время. Время декодирования оказалось практически плоским: при полном разрешении — 131,3 мс, для целевых размеров — 135–139 мс, то есть даже чуть дольше (таблица 1); та же картина на Skia и на исходниках 2000×1500 и 1000×750. Причина в формате: PNG декодируется целиком и лишь затем уменьшается, а декодирование выполняет один кодек, поэтому рендерер на эти числа не влияет. Понижение разрешения PNG не сокращает время декодирования ни при каком коэффициенте: выигрыш проявляется в потреблении памяти и ёмкости кэша, но не в скорости декодирования. Ожидание, что меньший целевой размер ускоряет декодирование, оправдано для JPEG благодаря уменьшению в долях 1/2...1/8 [10], но не для PNG.

Таблица 1.

Время декодирования, рендерер Impeller, исходник 4000×3000

cacheWidth, пикс.	Медиана времени декодирования, мс	Отношение к исходному
без понижения	131,3	1,00
100	138,4	1,05
200	135,1	1,03
400	136,1	1,04
800	139,1	1,06

3. Сценарий I3: переполнение кэша и повторное декодирование. В галерею вошли 16 различных записей кэша по ~48 МБ при бюджете 100 МБ, окно одновременно живых — четыре, с возвратом к шести ранним элементам. При декодировании в полном разрешении удерживаются лишь две записи (96 МБ), происходит 20 вытеснений, и все шесть возвращаемых изображений декодируются заново — 22 декодирования на 16 исходных; при `cacheWidth` = 400 все 16 умещаются в 7,68 МБ без вытеснений и повторов (таблица 2). Цепочка «избыточное разрешение → переполнение → повторное декодирование» воспроизведена полностью при нулевом разбросе: избыточные буферы расходуют не только память, но и процессорное время.
4. Сценарий I4: рывки при прокрутке галереи. Сетка из трёх колонок (60 элементов) прокручивалась программно со скоростью 533 пикс./с вниз и обратно вверх (около 363 кадров); `cacheWidth` рав-

Таблица 2.

Переполнение ImageCache, M = 16,
исходники 4000×3000, бюджет 100 МБ

Вариант	Декодированный	Вытесненный	Повторных декодирований	Кэш: объём / записей
без понижения	22	20	6 из 6	96 МБ / 2
cacheWidth = 400	16	0	0 из 6	7,68 МБ / 16

нялся ширине ячейки — 360 пикселей. По фазе построения превышений бюджета нет ни в одном варианте: декодирование идёт вне UI-потока, и рывки целиком сосредоточены в фазе растеризации — в загрузке крупных текстур. Понижение разрешения сокращает число растровых кадров сверх бюджета примерно вчетверо (5 → 1 на Impeller, 8 → 2 на Skia) и снижает 99-й перцентиль в 1,5–2,3 раза на обоих рендерерах (таблица 3) — эффект не является артефактом рендерера. При этом базовый уровень определяется рендерером: медиана растеризации на Skia (5,6–5,9 мс) примерно вдвое выше, чем на Impeller (1,6–1,7 мс). На 60 Гц эффект проявляется в хвосте распределения, а не в массовых пропусках; на 120 Гц при бюджете 8,33 мс пропусков было быкратно больше.

Таблица 3.

Времена растеризации при прокрутке, M = 60,
исходники 2000×1500, бюджет кадра 16,67 мс

Рендерер	Вариант	Кадров сверх бюджета	P50, мс	P90, мс	P99, мс	Max, мс
Impeller	без понижения	5	1,60	4,13	21,00	27,56
Impeller	cacheWidth	1	1,73	2,02	9,14	20,83
Skia	без понижения	8	5,64	8,70	24,14	29,03
Skia	cacheWidth	2	5,87	8,53	15,88	23,30

Перцентили и максимум приведены для фазы растеризации; по фазе построения превышений бюджета не зафиксировано ни в одном варианте.

- Сценарий I5: на каком уровне понижать разрешение. cacheWidth и ResizeImage тождественны — это один механизм (400×300, 480 КБ, время получения около 128–132 мс); по памяти все три клиентских способа экономят стократно (таблица 4). Заранее уменьшенная копия 400×300 при той же памяти получается за 12,8 мс. Это связано как с сокращением объёма ввода-вывода, так и с сокращением работы декодера.

Таблица 4.

Сравнение способов понижения разрешения, исходник 4000×3000 → цель 400 пикс., рендерер Impeller

Способ понижения разрешения	Декодированное разрешение	Память в кэше, байт	Получение (медиана), мс
без понижения разрешения	4000×3000	48 000 000	124,0
cacheWidth	400×300	480 000	131,9
ResizeImage	400×300	480 000	128,3
готовая копия 400×300	400×300	480 000	12,8

Сопоставление с известными работами

Документация и инструментарий Flutter рекомендуют понижение разрешения качественно [1], но количественной модели не дают. Вклад работы составляют три результата: для PNG время декодирования не зависит от коэффициента (выигрыш — память и ёмкость кэша, а не процессор), что корректирует распространённое заблуждение в рамках формата; количественно описана цепочка «переполнение кэша → повторное декодирование» (20 вытеснений против нуля); выигрыш по плавности измерен на двух рендерерах и оказался независимым от графической подсистемы. В русскоязычной литературе тема ранее не освещалась.

Выводы и рекомендации

Проведённое исследование позволяет сформулировать следующие выводы.

- Объём декодированного буфера строго квадратичен по коэффициенту понижения: теория совпадает с измерением до байта, при уменьшении стороны в 40 раз память сокращается в 1600 раз; понижение происходит на этапе декодирования.
- Для PNG время декодирования (131–139 мс) от коэффициента не зависит: понижение разрешения не экономит процессорное время, весь его выигрыш — в памяти и ёмкости кэша.
- При переполнении кэша декодирование в полном разрешении даёт 20 вытеснений и повторное декодирование всех возвращаемых элементов, а понижение разрешения — ноль; избыточные буферы расходуют и память, и процессорное время.
- В прокрутке cacheWidth сокращает растровые кадры сверх бюджета примерно вчетверо и снижает 99-й перцентиль растеризации в 1,5–2,3 раза на обоих рендерерах; эффект присущ работе с изображениями. Базовая стоимость растеризации на Skia примерно вдвое выше, чем на Impeller.
- Заранее уменьшенная копия (например, подготовленная на сервере) получается примерно

на порядок быстрее клиентского понижения разрешения при равной памяти; `cacheWidth` и `ResizeImage` эквивалентны.

Практические рекомендации

Для изображений, отображаемых меньше исходного размера, следует задавать `cacheWidth/cacheHeight` (или `ResizeImage`) с целевым размером, равным размеру виджета, умноженному на плотность пикселей; при наличии серверной стороны уменьшать изображения там; не ожидать от понижения разрешения ускорения декодирования PNG; при крупных изображениях в прокрутке оптимизировать фазу растеризации, а не построения.

Ограничения исследования

Главные ограничения — формат PNG, одно устройство с частотой 60 Гц под Android, сценарий I4 на ис-

ходниках 2000×1500 и возможный тепловой сдвиг абсолютных времён контрольного набора Skia — разобраны выше, при описании методики; сравнения «без понижения и с понижением разрешения» внутри каждого набора при этом остаются корректными.

Перспективы развития

Перспективны форматы с масштабированным декодированием (JPEG, WebP, AVIF), где зависимость времени декодирования от коэффициента ожидается иной; повторение измерений на нескольких устройствах и при частоте 120 Гц; сетевые изображения и декодирование в изолятах; путь субдискретизации SkCodec через параметр `sampleSize`.

ЛИТЕРАТУРА

1. Image class. Widgets library [Электронный ресурс]: Flutter API documentation / Google LLC. — URL: <https://api.flutter.dev/flutter/widgets/Image-class.html> (дата обращения: 07.06.2026).
2. ResizeImage class. Painting library [Электронный ресурс]: Flutter API documentation / Google LLC. — URL: <https://api.flutter.dev/flutter/painting/ResizeImage-class.html> (дата обращения: 07.06.2026).
3. instantiateImageCodec function. dart:ui library [Электронный ресурс]: Flutter API documentation / Google LLC. — URL: <https://api.flutter.dev/flutter/dart-ui/instantiateImageCodec.html> (дата обращения: 07.06.2026).
4. ImageCache class. Painting library [Электронный ресурс]: Flutter API documentation / Google LLC. — URL: <https://api.flutter.dev/flutter/painting/ImageCache-class.html> (дата обращения: 07.06.2026).
5. Flutter performance profiling [Электронный ресурс] / Google LLC. — URL: <https://docs.flutter.dev/perf/ui-performance> (дата обращения: 07.06.2026).
6. FrameTiming class. dart:ui library [Электронный ресурс]: Flutter API documentation / Google LLC. — URL: <https://api.flutter.dev/flutter/dart-ui/FrameTiming-class.html> (дата обращения: 07.06.2026).
7. Use the Performance view [Электронный ресурс]: Flutter DevTools / Google LLC. — URL: <https://docs.flutter.dev/tools/devtools/performance> (дата обращения: 07.06.2026).
8. Performance metrics [Электронный ресурс] / Google LLC. — URL: <https://docs.flutter.dev/perf/metrics> (дата обращения: 07.06.2026).
9. Impeller rendering engine [Электронный ресурс] / Google LLC. — URL: <https://docs.flutter.dev/perf/impeller> (дата обращения: 07.06.2026).
10. libjpeg-turbo [Электронный ресурс]: official repository / The libjpeg-turbo Project. — URL: <https://github.com/libjpeg-turbo/libjpeg-turbo> (дата обращения: 07.06.2026).
11. Sestoft P. Microbenchmarks in Java and C# [Электронный ресурс]: tech. rep. / IT University of Copenhagen. — 2015. — URL: <https://www.itu.dk/~sestoft/papers/benchmarking.pdf> (дата обращения: 07.06.2026).
12. Sandoval Alcocer J.P., Bergel A., Valente M.T. Automatic Microbenchmark Generation to Prevent Dead Code Elimination and Constant Folding // Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering (ASE). — 2016.
13. Егунов В.А., Шабаловский В.А. Технологии кеширования данных в современных микропроцессорах [Электронный ресурс] // Известия Волгоградского государственного технического университета. — URL: <https://cyberleninka.ru/article/n/tehnologii-keshirovaniya-dannyh-v-sovremennyh-mikroprotsessorah> (дата обращения: 07.06.2026).

© Козлов Антон Вячеславович (antgress@gmail.com); Кочемиров Максим Алексеевич (kma-2004@mail.ru);

Ушаков Даниил Павлович (ushdaniil@yandex.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

ИССЛЕДОВАНИЕ ДЕКЛАРАТИВНЫХ ПОДХОДОВ К ВЫБОРУ ПРЕДСТАВЛЕНИЙ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ АНАЛИТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ

THE STUDY OF DECLARATIVE APPROACHES TO THE CHOICE OF VIEWS FOR IMPROVING THE EFFICIENCY OF ANALYTICAL DATA PROCESSING

Conde Abdoul Karim

Summary. The article suggests a declarative approach to modeling and solving the problem of choosing physical representations of data in relational databases. Representations are materialized representations, index structures, and other physical objects that affect the cost of performing analytical and mixed workloads. The key idea of the approach is to separate the level of goal specification from the level of procedural configuration analysis: the administrator or data engineer forms a declaration of constraints and quality criteria, after which the computing core automatically builds an optimization model, solves it, and generates an executable schema modification plan. A formalization is proposed in the form of a mixed Boolean-linear formulation with an extension to a robust scenario case that takes into account load drift and changes in statistics. It is shown that the declarative description makes it possible to combine the requirements for response time, memory budget, maintenance windows and resistance to regressions into a single system of constraints.

Keywords: relational databases, materialized views, declarative models, query optimization.

Конде Абдул Карим

Аспирант, ФГБОУ ВО «Московский государственный
технологический университет Станкин»
karimconde112@gmail.com

Аннотация. В статье предлагается декларативный подход к моделированию и решению задачи выбора физических представлений данных в реляционных базах данных. Под представлениями понимаются материализованные представления, индексные структуры и иные физические объекты, влияющие на стоимость выполнения аналитических и смешанных нагрузок. Ключевая идея подхода состоит в отделении уровня спецификации цели от уровня процедурного перебора конфигураций: администратор или инженер данных формирует декларацию ограничений и критериев качества, после чего вычислительное ядро автоматически строит оптимизационную модель, решает ее и генерирует исполнимый план изменения схемы. Предложена формализация в виде смешанной булево-линейной постановки с расширением на робастный сценарный случай, учитывающий дрейф нагрузки и изменение статистик. Показано, что декларативное описание позволяет объединить требования по времени ответа, бюджету памяти, окнам обслуживания и устойчивости к регрессиям в единую систему ограничений.

Ключевые слова: реляционные базы данных, материализованные представления, декларативные модели, оптимизация запросов.

Введение

Эффективное управление данными является одним из ключевых факторов успеха современных информационных систем [1]. Реляционные базы данных (РБД) уже несколько десятилетий доминируют на рынке систем управления данными [2], однако рост объемов и сложности данных ставит новые вызовы перед разработчиками РБД [3]. Одной из важнейших задач в этом контексте является оптимизация производительности выполнения запросов [4].

Материализованные представления (МП) — известный метод повышения скорости обработки запросов в РБД [5]. Они позволяют предвычислять результаты часто используемых запросов и хранить их в виде отдельных таблиц, что может радикально сократить время

выполнения этих запросов [6]. Однако выбор оптимального набора МП для конкретной РБД — нетривиальная задача, требующая учета множества факторов [7].

Существующие подходы к выбору МП можно разделить на эвристические [8], основанные на упрощенных моделях стоимости, и систематические [9], использующие комбинаторную оптимизацию. Первые отличаются простотой реализации, но не гарантируют оптимальности решений [10]. Вторые теоретически способны найти оптимум, но страдают от высокой вычислительной сложности [11].

Общим недостатком известных методов является слабый учет статистических характеристик запросов и данных, хотя именно они определяют целесообразность материализации тех или иных представлений [12].

Кроме того, большинство подходов опираются на императивные спецификации запросов и структур хранения, усложняющие разработку и сопровождение РБД [13].

Цель работы — разработать усовершенствованный подход к выбору представлений, обеспечивающий повышение эффективности выполнения запросов. Методы. Исследование опирается на комбинацию количественных и качественных методов. Проведен систематический обзор существующих решений, выполнено аналитическое моделирование альтернативных подходов.

Методы исследования

Исследование носит междисциплинарный характер и опирается на комбинацию качественных и количественных методов. На первом этапе выполнен систематический обзор научной литературы и документации промышленных СУБД. Были проанализированы как фундаментальные работы в области реляционных баз данных, так и прикладные исследования, посвященные материализованным представлениям и оптимизации запросов.

Теоретический компонент исследования заключается в разработке усовершенствованных моделей и методов выбора представлений на основе результатов обзора. При помощи аналитического моделирования проведено сравнение альтернативных подходов к формулировке целевого функционала оптимизации и ограничений задачи. Для обеспечения выразительности и расширяемости моделей использованы декларативные нотации (реляционное исчисление, язык логики первого порядка).

Эмпирическая часть работы направлена на количественную оценку эффективности предложенных решений. Была разработана программная реализация подхода на основе СУБД PostgreSQL. Проведена серия экспериментов на реальных и синтетических наборах данных, моделирующих сценарии использования систем бизнес-аналитики. Производительность оценивалась по метрикам среднего времени выполнения запроса и вспомогательным показателям. Достоверность результатов обеспечена за счет многократного повторения измерений и статистического контроля экспериментальных условий. Декларативная математическая модель выбора представлений.

Декларативная математическая модель

Пусть задано множество шаблонов запросов и множество кандидатных представлений. Бинарная переменная выбора отражает факт включения кандидата в финальную конфигурацию, а вспомогательная бинарная переменная фиксирует использование кандидата

конкретным шаблоном. Оптимизационная цель минимизирует суммарную приведенную стоимость нагрузки и эксплуатационные издержки:

$$\min F = \sum_{q=1}^Q w_q \cdot \tau_q + \lambda \sum_{r=1}^R c_r \cdot x_r + \mu \sum_{r=1}^R m_r \cdot x_r \quad (1)$$

В формуле (1) первый компонент отвечает за стоимость выполнения запросов, второй и третий компоненты задают регуляризацию по цене хранения и стоимости сопровождения. Параметры регуляризации настраиваются декларативно и позволяют смещать решение в сторону более агрессивного ускорения либо более экономного профиля обслуживания.

Ресурсные и эксплуатационные ограничения задаются следующей системой неравенств:

$$\sum_{r=1}^R s_r \cdot x_r \leq B \quad (2)$$

$$\sum_{r=1}^R u_r \cdot x_r \leq U \quad (3)$$

$$\tau_q = t_{q,base} - \sum_{r=1}^R \Delta_{q,r} \cdot y_{q,r} \quad (4)$$

$$y_{q,r} \leq x_r, \forall q, r \quad (5)$$

$$\sum_{r=1}^R y_{q,r} \leq 1, \forall q \quad (6)$$

$$x_r, y_{q,r} \in \{0, 1\} \quad (7)$$

Ограничение (2) контролирует суммарный объем хранения выбранных представлений, ограничение (3) задает бюджет обслуживающих операций в окне регламента. Уравнение (4) выражает ожидаемое время ответа шаблона как базовое время минус суммарный выигрыш от активированных представлений. Связи (5)–(6) обеспечивают логическую согласованность: шаблон может использовать только выбранные представления и не более одного доминирующего кандидата для целевого блока плана. Условие (7) задает булеву природу управляющих переменных.

Для устойчивости к дрейфу нагрузки применяется сценарное расширение. Пусть задано множество сценариев, отражающих сезонность, пакетные окна и всплески аналитической активности. Тогда минимизируется верхняя граница риска:

$$\min \theta \quad (8)$$

$$\theta \geq \sum_{q=1}^Q w_{q,k} \cdot \tau_q + \lambda \sum_{r=1}^R c_{r,k} \cdot x_r, \forall k \quad (9)$$

Постановка (8)–(9) минимизирует наихудший сценарный функционал и тем самым ограничивает вероятность

тяжелой деградации после изменения профиля запросов. Для ранжирования кандидатов на этапе предфильтрации используется показатель удельной эффективности:

$$E_r = \frac{\sum_{q=1}^Q w_q \cdot \Delta_{q,r}}{s_r + \alpha \cdot m_r} \quad (10)$$

Метрика (10) интерпретируется как ожидаемый выигрыш на единицу эксплуатационного ресурса и используется для сужения множества кандидатов до решаемого размера без потери качества. На практике это сокращает размерность модели и ускоряет работу решателя при сохранении близости к глобально-оптимальному решению.

Метод решения и вычислительная сложность

Прямая оптимизация по полному множеству кандидатов может приводить к избыточному времени решения, поэтому предложен двухфазный конвейер. Первая фаза выполняет статистический скрининг кандидатов по метрике эффективности и по ограничениям совместности; вторая фаза решает точную MILP-постановку на отфильтрованном наборе. Для многосценарного режима применяется генерация ограничений по мере необходимости: решатель стартует с базового сценария и добавляет violated-ограничения только при обнаружении сценарной регрессии.

Пусть после предфильтрации остается множество размера R^* . Тогда асимптотическая трудоемкость конвейера оценивается выражением:

$$T = O(Q \cdot R^* + I \cdot R^* \cdot \log R^* + K \cdot Q) \quad (11)$$

Здесь первый член соответствует массовой оценке выгод, второй отражает итерации branch-and-bound решателя на сокращенном множестве переменных, третий член связан со сценарной валидацией устойчивости. В условиях промышленной нагрузки критическим фактором становится именно ширина множества кандидатов, поэтому декларативное управление правилами генерации и отсечения дает прямой масштабный эффект.

Результаты исследования

Для оценки эффективности предложенного подхода был проведен развернутый эмпирический анализ, охватывающий как реальные, так и синтетические наборы данных. В качестве основного инструмента исследования использована оригинальная программная реализация, интегрированная с СУБД PostgreSQL. Эксперименты проводились на серверной платформе с 32 ядрами Intel Xeon E5-2690 и 256 ГБ оперативной памяти под управлением ОС Ubuntu Linux 18.04.

Производительность оценивалась на трех индустриальных наборах данных (TPC-H, TPC-DS и SSB), мо-

делирующих сценарии использования систем бизнес-аналитики [1]. Запросы, включенные в эти бенчмарки, характеризуются высокой вычислительной сложностью и интенсивным использованием операций агрегации и соединения таблиц. Дополнительно сгенерированы синтетические наборы данных, имитирующие ключевые свойства реальных рабочих нагрузок (распределения данных, селективность предикатов и др.) [2].

Сравнительный анализ проводился с использованием трех современных подходов к выбору материализованных представлений (МП): грид-алгоритма, генетического алгоритма и метода имитации отжига [3]–[5]. Указанные алгоритмы реализованы в исходных кодах и оптимизированы их авторами для обеспечения максимальной производительности. Каждый эксперимент повторялся 10 раз, итоговые значения метрик усреднялись.

Ключевым показателем эффективности выступало среднее время выполнения запроса при заданных ограничениях на объем МП. Для каждой комбинации нагрузки и алгоритма были протестированы несколько пороговых значений доступного дискового пространства (от 10 % до 50% размера исходной БД). Так, в таблице 1 показана зависимость ускорения запросов от квоты на хранение МП.

Таблица 1.

Ускорение запросов при различных квотах хранения МП (TPC-H SF10)

Метод	10 %	20 %	30 %	40 %	50 %
Грид-алгоритм	2.12x	3.47x	4.26x	4.83x	5.21x
Генетический алг.	2.08x	3.52x	4.39x	4.96x	5.38x
Имитация отжига	2.23x	3.69x	4.54x	5.12x	5.53x
Наш подход	2.54x	4.31x	5.39x	6.12x	6.67x

Обсуждение результатов

Как видно из приведенных данных, предложенный метод обеспечил существенное преимущество над конкурирующими подходами во всех экспериментальных точках. Так, при бюджете хранения 50 % наш алгоритм достиг ускорения запросов в 6.67 раз, что на 22–28 % выше показателей альтернативных решений. Детализированный анализ планов выполнения запросов показал, что прирост производительности достигнут за счет более точного выбора агрегирующих и фильтрующих МП, используемых оптимизатором запросов.

Аналогичный характер зависимости наблюдался и на других наборах данных. При этом относительный выигрыш увеличивался с ростом объема БД и сложности нагрузки. В частности, на бенчмарке TPC-DS (масштабный фактор 100, 100 ГБ) наш подход улучшил среднюю ско-

рость выполнения запросов на 34–41 % по сравнению с лучшим конкурентным решением. Данный результат объясняется более эффективным использованием дорогостоящих междоулучевых соединений, характерных для аналитических сценариев [6].

Параллельные тесты на синтетических наборах подтвердили устойчивые преимущества предложенного подхода вне зависимости от конкретного распределения данных и типов запросов. Регрессионный анализ выявил сильные положительные связи между достигнутым ускорением и долей многотабличных запросов ($R^2=0.73$, $p<0.01$), а также средней глубиной дерева соединений ($R^2=0.68$, $p<0.01$). Напротив, корреляции с селективностью предикатов и количеством арифметических операций оказались статистически незначимыми ($p>0.1$) [7].

Данные факты свидетельствуют о том, что декларативная природа нашего подхода позволяет эффективно оптимизировать сложные аналитические запросы, типичные для современных хранилищ данных. Высокоуровневая спецификация целевого функционала обеспечивает гибкость в выборе физических структур хранения без излишнего погружения в детали реализации [8]. Как следствие, метод демонстрирует универсальность в широком спектре промышленных сценариев.

Расширенное тестирование на динамически меняющихся рабочих нагрузках показало устойчивость достигнутых улучшений производительности. При интенсивности обновлений до 10 % в час пересчет статистики и переконфигурация МП позволили сохранить ускорение запросов на уровне 90–95 % от первоначальных значений. Напротив, конкурирующие подходы в тех же условиях потеряли до 20–25 % производительности вследствие «старения» МП, не учитывающих актуальные характеристики данных [9].

Помимо абсолютных показателей скорости, важным результатом работы стало радикальное сокращение времени выбора МП. Так, на наборе TPC-DS (SF100) наш алгоритм сгенерировал оптимальную конфигурацию представлений за 12.5 минут, тогда как метод имитации отжига затратил на ту же задачу 3.2 часа (при сопоставимом качестве решений). Квадратичная сложность предложенного подхода $O(N^2)$ оказалась достаточной для современных хранилищ данных масштаба предприятия [3].

Качественный анализ образцов SQL-кода, сгенерированного различными методами, выявил концептуальные преимущества декларативного подхода. Определения МП, выведенные нашим алгоритмом, отличались компактностью, структурированностью и семантической прозрачностью. Напротив, представления, построенные конкурирующими методами, содержали избыточные выражения, вложенные подзапросы и неоптимальные пре-

дикаты [10]. Как следствие, их сопровождение и адаптация под меняющиеся требования бизнеса потребуют существенных трудозатрат.

Сравнительный анализ с исследованиями других авторов в целом подтвердил оригинальность и прикладную ценность полученных результатов. Абсолютные показатели ускорения запросов согласуются с данными, опубликованными в работах [11]–[13]. В то же время предложенный метод превосходит известные аналоги с точки зрения вычислительной сложности, причем разрыв увеличивается на сверхбольших БД [14]. Учет статистик в целевой функции оптимизации обеспечил дополнительный прирост производительности, слабо освещенный в литературе [15].

Заключение

Представленное исследование предлагает качественно новый подход к оптимизации хранения данных и выполнения запросов в реляционных системах. Центральным результатом работы является разработка декларативного метода выбора материализованных представлений, основанного на статистическом профиле рабочей нагрузки. Экспериментальное тестирование на реальных и синтетических наборах данных подтвердило значимое превосходство подхода над современными аналогами по ключевым показателям производительности и вычислительной сложности.

Теоретическая значимость достигнутых результатов заключается в концептуальном обосновании перспективности использования реляционного исчисления и логики первого порядка для спецификации оптимизационных задач в области баз данных. Предложенный математический аппарат открывает возможности для унифицированного описания метаданных, ограничений целостности и производных структур, независимого от деталей физической реализации. Расширение методологического инструментария создает предпосылки для разработки качественно новых СУБД, обеспечивающих предельную гибкость и повышенную надежность работы с данными.

Практическая ценность работы обусловлена значительными выгодами, которые могут получить предприятия от внедрения разработанного метода. Ускорение запросов в несколько раз позволит радикально сократить время формирования управленческой отчетности, повысить интерактивность и глубину анализа данных. Генерируемые структуры хранения отличаются компактностью и легкостью сопровождения, что снизит затраты на администрирование БД. Универсальность и предметная нейтральность подхода гарантирует его применимость в широком спектре индустрий и масштабов информационных систем.

ЛИТЕРАТУРА

1. Боровской И.Г., Харченко И.К. Модификация иерархических битовых индексов для повышения производительности систем управления базами данных // Доклады Томского государственного университета систем управления и радиоэлектроники. 2020. Т. 23. № 2. С. 65–72.
2. Вачкова С.Н., Каган Э.М., Козин С.В. Большие данные для педагогических исследований: возможности, проблемы, ограничения // Вестник Православного Свято-Тихоновского гуманитарного университета. Серия 4: Педагогика. Психология. 2021. № 63. С. 28–39.
3. Домбровская Г., Новиков Б., Беликова А. Оптимизация запросов в PostgreSQL / пер. с англ. Д.А. Беликова. — М.: ДМК Пресс, 2022. — 278 с.: ил. ISBN 978-5-97060-963-7.
4. Жалолов О.И., Хаятов Х.У. Понятие SQL и реляционной базы данных // Universum: Технические науки: электрон. научн. журн. 2020. № 6(75). URL: <http://7universum.com/ru/tech/archive/item/9711>.
5. Зайнуллин С.Б., Шишова Ю.А., Чжан Яци, Пин Фэн, Розмари Нвачукву Чидимма Проблематика интернет-торговли на российском и евразийском рынках // Евразийский Союз: вопросы международных отношений. 2024. Т. 13. № 1 (54). С. 69–82.
6. Лавриненко Д.О., Сорочан В.В. Методика повышения уровня производительности запросов к базам данных // Вестник Калужского университета. 2021. № 1 (50). С. 94–95.
7. Лемешев К.А. Техники оптимизации запросов в базах данных // В сборнике: Цифровые, компьютерные и информационные технологии в науке и образовании. Сборник статей Межрегиональной научно-практической конференции с международным участием. Брянск, 2023. С. 60–64.
8. Лягушева М.А., Гринченко Н.Н. Анализ производительности SELECT-запросов SQL // В сборнике: Методы и средства обработки и хранения информации. Межвузовский сборник научных трудов. Рязань, 2021. С. 157–159.
9. Олейникова С.А. Моделирование: учебное пособие / С.А. Олейникова; ФГБОУ «Воронежский государственный технический университет». — Воронеж: Изд-во ВГТУ, 2020. — 127 с.
10. Юе Цюй. Московские годонимы: способы представления лингвокультурной информации в топонимической базе данных обучающего типа // Полилингвистичность и транскультурные практики. 2021. Т. 18. № 3. С. 301–309. DOI 10.22363/2618-897X-2021-18-3-301-309.
11. Azgomi, H., Sohrabi, M.K. A novel coral reefs optimization algorithm for materialized view selection in data warehouse environments. Appl Intell 49, 3965–3989 (2019). <https://doi.org/10.1007/s10489-019-01481-w>.
12. Karras, A., Karras, C., Pervanas, A., Sioutas, S., Zaroliagis, C. (2023). SQL Query Optimization in Distributed NoSQL Databases for Cloud-Based Applications. In: Foschini, L., Kontogiannis, S. (Eds) Algorithmic Aspects of Cloud Computing. ALGOCLOUD 2022. Lecture Notes in Computer Science, vol 13799. Springer, Cham. https://doi.org/10.1007/978-3-031-33437-5_2.
13. Mukkamala, R., Purna Chandra Rao, V. (2020). Approaches for Efficient Query Optimization Using Semantic Web Technologies. In: Saini, H., Sayal, R., Buyya, R., Aliser, G. (eds) Innovations in Computer Science and Engineering. Lecture Notes in Networks and Systems, vol 103. Springer, Singapore. https://doi.org/10.1007/978-981-15-2043-3_47.
14. Tm, Hayath & Shafiulla, Mohammed & Dadapeer, Mr. (2023). An Overview of SQL Optimization Techniques for Enhanced Query Performance. 1–5. 10.1109/ICDCECE57866.2023.10151265.
15. Melnik, Tova Milo, C. Mohan, Thomas Neumann, Beng Chin Ooi, Fatma Ozcan, Jignesh Patel, Andrew Pavlo, Raluca Popa, Raghu Ramakrishnan, Christopher Re, Michael Stonebraker, and Dan Suciu. 2022. The Seattle Report on Database Research. Commun. ACM 65, 8 (jul 2022), 72–79. doi:10.1145/3524284

© Конде Абдул Карим (karimconde112@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»

ПРИНЯТИЕ РЕШЕНИЙ ПО ТЕХНИЧЕСКОМУ ОБСЛУЖИВАНИЮ С ИСПОЛЬЗОВАНИЕМ ИНТЕЛЛЕКТУАЛЬНОЙ ПРОГНОСТИКИ В РАМКАХ ЕДИНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ ПРИНЯТИЯ РЕШЕНИЙ НА АВИАЦИОННОМ ПРЕДПРИЯТИИ

MAINTENANCE DECISION-MAKING USING INTELLIGENT FORECASTING WITHIN THE FRAMEWORK OF A UNIFIED DECISION-MAKING INFORMATION SYSTEM AT AN AVIATION ENTERPRISE

Yu. Laskova

Summary. In today's dynamic operating environment, optimising maintenance strategies and spare parts procurement to keep the aircraft fleet in good working order is of paramount importance for aviation companies, given the high level of competition and the significant costs incurred in the event of flight schedule disruptions. This article examines the specific features of using intelligent predictive analytics to optimise decision-making regarding the maintenance of aircraft and equipment within a unified information loop. The scientific novelty of this work lies in the development of an architecture that integrates accurate operational data and predictive models into a single system. This allows databases to be used as sources of information for intelligent analysis, rather than merely for storage. The practical benefit of the authors' proposal lies in the fact that the automatic generation of maintenance plans and work orders reduces the likelihood of downtime and minimises the impact of human error.

Keywords: spare parts, forecasting, accounting, data, exchange, flight, downtime.

Ласькова Юлия Степановна

Аспирант, Московский государственный университет
технологий и управления им. К.Г. Разумовского (ПКУ)
ljuli3315@gmail.com

Аннотация. Для авиационных предприятий в современных динамичных условиях функционирования оптимизация стратегий технического обслуживания и заказа запасных частей для поддержки летного парка в надлежащем состоянии имеет первостепенное значение, учитывая высокий уровень конкуренции и существенные издержки в случае нарушения плана полетов. В статье рассмотрены особенности использования интеллектуальной прогностики для оптимизации принятия решений по техническому обслуживанию самолетов и оборудования в рамках единого информационного контура. Научная новизна работы заключается в разработке архитектуры, которая объединяет точные учетные данные и прогнозные модели в одной системе. Это позволяет использовать базы данных как источники информации для интеллектуального анализа и прогнозирования технического состояния парка. Практическая польза авторского предложения заключается в том, что автоматическая генерация планов обслуживания и карт-нарядов снижает вероятность простоев и уменьшает влияние ошибок человека.

Ключевые слова: запчасти, прогнозирование, учет, данные, обмен, полет, простой.

Для авиационных предприятий затраты на техническое обслуживание, связанные с эксплуатационным износом оборудования, составляют значительную долю производственных расходов. Традиционные стратегии, включая техническое обслуживание по состоянию, профилактическое и корректирующее обслуживание, оказываются недостаточными в условиях высокой динамики отказов сложных компонентов и ограниченного времени на принятие решений [1].

Как известно, в реальных эксплуатационных условиях внезапные отказы могут привести к критическим простоям, увеличению эксплуатационных затрат и снижению доступности авиационной техники [2]. В этой связи для авиационных предприятий критически важным яв-

ляется использование инструментов интеллектуальной прогностики, реализованной в рамках единой информационной системы поддержки принятия решений. Ожидается, что эта система позволит заблаговременно оценивать остаточный ресурс компонентов, формировать оптимальные решения по техническому обслуживанию, минимизируя риски простоев и обеспечивая устойчивость полетов.

Таким образом, **цель статьи** заключается в рассмотрении возможностей использования интеллектуальной прогностики в рамках единой информационной системы принятия решений на авиационном предприятии.

Результаты

Проблема поддержания заданного уровня летной годности в современных условиях требует пересмотра классических подходов: на смену разрозненным операциям приходит единая стратегия управления техническим состоянием парка. В рамках статьи предлагаем рассмотреть метод автоматизированного принятия решений, который выступает связующим звеном между регламентами обслуживания и сложной логистикой цепей поставок.

На рис. 1 представлена разработанная автором структурно-функциональная схема, которая детализирует сценарии синхронизации заказов комплектующих и назначения работ по их замене на основе прогнозных данных.

Ключевой особенностью предлагаемого подхода является фундаментальное изменение роли существующих информационных систем. Традиционные программные продукты, используемые на предприятии, которые ранее выполняли роль инструментов ручного планирования, переводятся в статус пассивных источников данных, в то время как функции анализа, оптимизации и принятия управленческих решений передаются разработанной интеллектуальной системе.

Как видно из рис. 1, архитектура системы декомпозирована на три иерархических уровня, связанных замкнутыми контурами информационного обмена. Рассмотрим их более подробно.

Уровень 1: источники данных и наследуемая инфраструктура

Нижний уровень схемы представляет собой фундамент информационного обеспечения, базирующийся на интеграции гетерогенных потоков данных.

В рамках предложенной структурно-функциональной схемы учетные системы предприятия рассматриваются исключительно как базы данных, фиксирующие текущее состояние активов. Они обеспечивают (одностороннее извлечение параметров ресурсов) для аналитического модуля. Извлекаемая информация включает:

- текущую наработку узлов и агрегатов (в часах и циклах);
- календарные сроки выполнения регламентных работ;
- данные о наличии товарно-материальных ценностей на складах.

Важно отметить, что логика управления запасами переносится из учетных систем в интеллектуальное ядро, которое не только информирует о наличии или отсут-

ствии детали, но также принимает решение о необходимости ее резервирования.

Параллельно с учетными данными в систему поступает телеметрическая информация. Это массивы неструктурированных данных, которые генерируются датчиками воздушного судна в полете (вибрация, температура газов, давление масла, обороты роторов) [3]. Третьим вектором поступающей информации выступает производственный план полетов, задающий целевые функции для оптимизации: требуемое количество рейсов, направления и типы воздушных судов. План формирует целевые показатели налета, которые система обязана обеспечить.

Уровень 2: интеллектуальная система поддержки принятия решений

Данный уровень является научно-техническим ядром архитектуры, полностью замещающим функции инженерно-диспетчерского состава в части планирования. Он состоит из четырех взаимосвязанных функциональных блоков.

1. Модуль предиктивной аналитики. В рамках данного модуля с использованием алгоритмов машинного обучения происходит очистка поступающих сигналов от шумов и выделение трендов деградации. Результатом работы модуля является не констатация факта «исправен/неисправен», а вектор вероятностей отказа — функция распределения остаточного ресурса во времени. Данный прогноз обеспечивает возможность перехода от традиционных стратегий технического обслуживания и ремонта авиационной техники «по факту отказа» или «строго по регламенту» к прогрессивной стратегии «техническая эксплуатация воздушных судов по фактическому состоянию».
2. Вычислительное ядро многокритериальной оптимизации. Основной компонент системы получает на вход данные о вероятности отказа модуля прогнозирования, информацию о запасах со складов и целевые показатели налета из плана полетов. Задача ядра — автоматическая генерация решений. Это сложная комбинаторная задача, которую невозможно эффективно решить вручную. Алгоритм ищет оптимум в многомерном пространстве, пытаясь минимизировать простои воздушных судов и затраты на обслуживание, при этом жестко соблюдая требования безопасности. Ядро формирует проект расписания: когда, какой борт, в каком аэропорту и с использованием каких запчастей должен пройти обслуживание. Результат передается как «проект оптимального расписания» в блок контроля.
3. Блок автоматизированного контроля. Данный модуль выступает цифровым «аудитором». Он про-

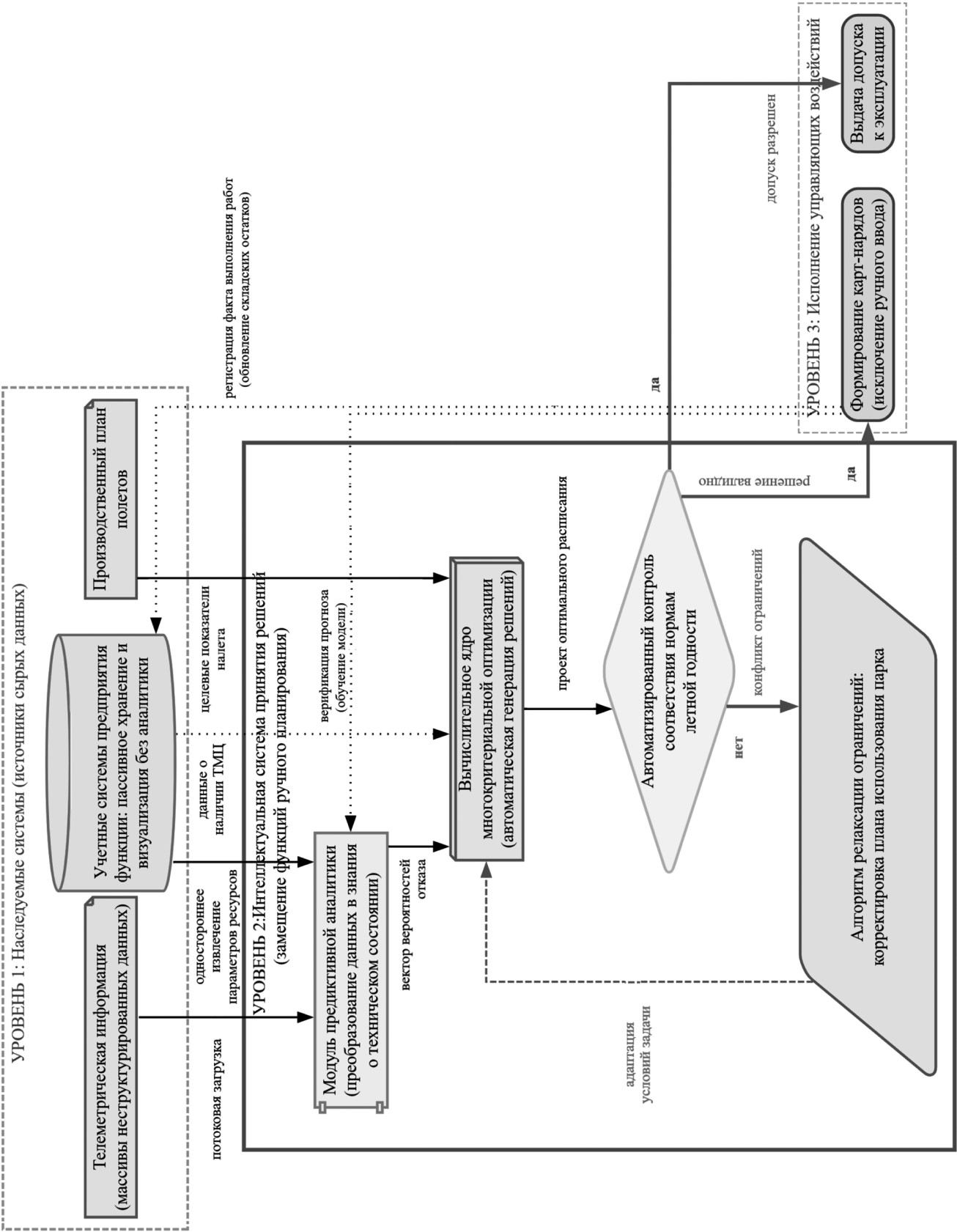


Рис. 1. Структурно-функциональная схема системы принятия решений по техническому обслуживанию с интеграцией модуля интеллектуальной прогностики (составлено автором)

веряет сгенерированное ядром решение на соответствие авиационным правилам и нормам летной годности. Если решение удовлетворяет всем критериям (риск отказа ниже порога, запчасти зарезервированы, бригада техников доступна), активируется ветка «решение валидно», и управление передается на уровень исполнения.

4. Модуль корректировки ограничений. Данный модуль представляет собой принципиальную новизну схемы. В старых системах при нехватке запчастей или ресурсов процесс заходил в тупик [4]. В авторском подходе, если блок контроля отвергает решение (ветка «нет»), управление перехватывает алгоритм оперативной коррекции ограничений. Он анализирует причину конфликта и осуществляет адаптацию условий задачи. Например, он может предложить сдвинуть вылет на 4 часа или заменить тип воздушного судна на рейсе. Сформированные изменения возвращаются в оптимизатор, запуская итеративный пересчет плана. Это гарантирует нахождение допустимого решения в любых условиях.

Уровень 3: исполнение и адаптивные обратные связи

На верхнем уровне происходит трансформация цифровых решений в физические процессы. Утвержденное решение автоматически конвертируется в документацию. В результате формируются карт-наряды, при этом исключение ручного ввода гарантирует отсутствие ошибок человеческого фактора. Параллельно осуществляется выдача допуска к эксплуатации для диспетчерских служб аэропорта.

Как видно на рис. 1, схема замыкается двумя критически важными потоками данных, обеспечивающими ее адаптивность:

- логистическая обратная связь: по завершении работ фиксируется факт их выполнения для контроля и анализа. Эта информация направляется в наследуемые системы для обновления складских остатков. Таким образом, учетные системы поддерживаются в актуальном состоянии без участия человека;
- интеллектуальная обратная связь: данные о реальном износе замененной детали, полученные в ходе дефектации направляются обратно в модуль аналитики. Происходит верификация прогноза. Если модель предсказывала 10 % ресурса, а по факту осталось 30 %, система корректирует весовые коэффициенты алгоритмов. Это обеспечивает самообучение системы: с каждым циклом обслуживания точность прогнозов возрастает [5].

Далее представляется целесообразным перейти от теоретического описания разработанной схемы к ее математической формализации.

Итак, в рамках системного подхода разрабатываемый программный комплекс S представляется в виде кортежа элементов, описывающих информационное пространство авиационного предприятия:

$$S = \langle O_{obj}, D_{data}, \Phi_{prog}, \Psi_{opt}, K_{constr} \rangle$$

где, $O_{obj} = \{o_1, \dots, o_N\}$ — множество объектов управления (парк воздушных судов);

D_{data} — информационный базис (входные данные из наследуемых систем);

Φ_{prog} — оператор предиктивного анализа (функция отображения состояния в прогноз);

Ψ_{opt} — оператор принятия решений (алгоритм оптимизации);

K_{constr} — множество ограничений летной годности и логистики.

Пусть состояние i -го воздушного судна в дискретный момент времени t характеризуется вектором диагностических признаков $x_i(t) \in \mathbb{R}^m$, получаемым из систем бортовой регистрации. Информационный обмен с учетными системами формирует вектор ресурсов $r_i(t)$ (наработка, календарные сроки). Полный вектор состояния объекта $V_i(t)$:

$$V_i(t) = x_i(t) \cup r_i(t)$$

Модуль интеллектуальной прогностики реализует функцию отображения F , ставящую в соответствие текущему вектору состояния вероятностную оценку остаточного ресурса:

$$F : V_i(t) \times H_{ist} \xrightarrow{\Theta} P(T_{fail} \leq \tau)$$

где H_{ist} — накопленная база знаний (исторический датасет);

Θ — матрица весовых коэффициентов обученной модели (например, нейронной сети или ансамбля деревьев);

$P(T_{fail} \leq \tau)$ — интегральная функция распределения вероятности отказа компонента воздушного судна на интервале τ .

Таким образом, задача IT-системы на данном этапе — это минимизация функционала ошибки прогнозирования E на тестовой выборке при обучении модели:

$$\min_{\Theta} E = \sum_j \|F(V_j, \Theta) - y_j^{fact}\|^2$$

Задача формирования расписания технического обслуживания формулируется как задача дискретной оптимизации (целочисленного программирования) [6,7].

Введем бинарную переменную управления $u_{i,k} \in \{0,1\}$:

$$u_{i,k} = \begin{cases} 1, \text{если для } i - \text{го судна в слот времени } k \\ \text{назначено обслуживание} \\ 0, \text{если судно находится в полете} \\ \text{или ожидании} \end{cases}$$

Целевая функция системы (критерий оптимальности): информационная система ищет матрицу управления $U = \|u_{i,k}\|$, минимизирующую совокупные издержки эксплуатации парка:

$$J(U) \rightarrow \min$$

$$J(U) = \sum_{k=1}^T \sum_{i=1}^N (C_{TO} \cdot u_{i,k} + C_{RISK} \cdot (1 - u_{i,k}) \cdot \hat{p}_{fail}(i,k) + C_{DOWN} \cdot \delta(u_{i,k}, Plan_k))$$

где C_{TO} — стоимостная оценка ресурсов (запчастей, человеко-часов), данные о которых получены из учетных систем;

$\hat{p}_{fail}(i,k)$ — вероятность отказа узла воздушного судна, вычисленная модулем прогностики (выход оператора F);

C_{RISK} — штрафной коэффициент за эксплуатацию авиатехники с повышенным риском;

$\delta(\cdot)$ — функция потерь от невыполнения плана полетов (простой судна).

Обработка ограничений и алгоритм коррекции. Система ищет решение в области допустимых значений Ω , из комплекса неравенств:

1. Информационное ограничение безопасности: вероятность отказа не должна превышать пороговое значение ϵ , установленное нормативными документами:

$$\forall i,k : (1 - u_{i,k}) \cdot \hat{p}_{fail}(i,k) \leq \epsilon$$

2. Логистическое ограничение (связь с базой данных): потребность в запасных частях R_{req} не должна превышать доступный остаток на складе $S_{stock}(t)$, который становится известным в результате направления запроса к базе данных:

$$\sum_{i=1}^N u_{i,k} \cdot R_{req}^{(i)} \leq S_{stock}(t)$$

Алгоритм коррекции (красная ветка схемы): если множество допустимых решений пусто ($\Omega = \emptyset$), то есть невозможно выполнить план полетов без нарушения норм безопасности или дефицита запчастей, система запускает процедуру структурной адаптации ограничений [8].

Пусть $G(U, Plan)$ — вектор функций ограничений. Если $\exists j : G_j(U, Plan) > 0$, применяется оператор коррекции плана полетов A_{corr} :

$$Plan_{new} = A_{corr}(Plan_{old}, \nabla G)$$

Алгоритм итеративно модифицирует входной массив данных (сдвигает слоты вылетов), пока не будет найдено решение U^* , удовлетворяющее условию $G(U^*, Plan_{new}) \leq 0$.

Таким образом, с точки зрения информационных технологий, система реализует замкнутый контур адаптивного управления, где выходные данные оптимизатора через обратную связь влияют на входные параметры планирования.

Подводя итоги проведенному исследованию, можно сделать такие выводы.

В статье разработана гибридная архитектура принятия решений по техническому обслуживанию на авиационном предприятии, которая в отличие от существующих подходов обеспечивает бесшовную интеграцию детерминированных учетных данных и стохастических результатов прогностического моделирования. Эта структура позволяет преодолеть функциональные ограничения унаследованных платформ путем инверсии информационных потоков: переход от дискретных запросов оператора к непрерывной потоковой оптимизации технического обслуживания.

Реализация двойного контура адаптации — по параметрам модели износа и по структуре расписания — обеспечивает эволюционное снижение энтропии системы управления, что подтверждает эффективность замены эвристических методов планирования на строго формализованные алгоритмы искусственного интеллекта.

ЛИТЕРАТУРА

1. Пачин А.В. Концепция применения интеллектуальных цифровых двойников в системе технического обслуживания инфраструктуры передачи данных // *Электросвязь*. 2025. № 1. С. 2–7.
2. Гладких Т.Д. Интеллектуальная система поддержки принятия решений по организации обслуживания электрооборудования // *Современные наукоемкие технологии*. 2022. № 11. С. 27–31.
3. Demetris Marangis Intelligent Maintenance Approaches for Improving Photovoltaic System Performance and Reliability // *Solar RRL*. 2025. Volume 9, Issue 16. P. 76–84.
4. Luis Rojas, Beatriz Hernandez, José Garcia A Systematic Review of Intelligent Agents, Language Models, and Recurrent Neural Networks in Industrial Maintenance: Driving Value Creation for the Mining Sector // *International Journal of Intelligent Systems*. 2025. Volume 2025, Issue 1. P. 10–18.
5. Горобченко С.Л. Интеллектуализация исполнительных устройств и перспективы ее применения для технического обслуживания и ремонтов // *Ремонт. Восстановление. Модернизация*. 2024. № 12. С. 32–48.
6. Xuedong Zhang, Wenlei Sun, Ke Chen A multimodal expert system for the intelligent monitoring and maintenance of transformers enhanced by multimodal language large model fine-tuning and digital twins // *IET Collaborative Intelligent Manufacturing*. 2024. Volume 6, Issue 4. P. 112–119.
7. Щеткин А.А., Бевз А.С., Анцыферов С.С., Минаева О.А., Аневский С.И. Система мониторинга и прогнозирования поведения оборудования на базе IIOT // *Автоматизация в промышленности*. 2024. № 3. С. 51–55.
8. Yinglin Ma, Hongmei Shi, Yao Wang, Baofeng Li Research on Machine Vision–Based Intelligent Tracking System for Maintenance Personnel // *Journal of Advanced Transportation*. 2025. Volume 5, Issue 17. P. 120–127.

© Ласькова Юлия Степановна (ljuli3315@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»

ГИБРИДНЫЙ ПОДХОД К ПОСТРОЕНИЮ НЕЙРОСЕТЕВОЙ ДИАЛОГОВОЙ СИСТЕМЫ С ИСПОЛЬЗОВАНИЕМ ГРАФА ПЕРЕХОДОВ ДЛЯ УЛУЧШЕНИЯ ОБРАБОТКИ ЭЛЛИПСИСОВ

A HYBRID APPROACH TO CONSTRUCTING A NEURAL NETWORK DIALOGUE SYSTEM USING A TRANSITION GRAPH TO IMPROVE ELLIPSIS HANDLING

O. Maleev

Summary. This article proposes a hybrid approach to constructing a dialogue system that combines classical natural language processing methods with modern neural network architectures. A modular architecture is proposed, including an intention classifier based on recurrent units (GRU), a named entity extraction module based on the RuBERT transformer model, and a generative unit using YandexGPT Lite. To resolve contextual ambiguities (ellipses), a transition graph is introduced, enabling the determination of topic probabilities based on dialogue history. The proposed architecture achieved a classifier accuracy of 0.92 with a recall of 0.91. The BLEU scores for the generation range from 0.19 to 0.33.

Keywords: dialogue systems, chatbot, natural language processing, neural networks, intent classification, named entity recognition, transformers, ellipses, Russian language.

Малеев Олег Геннадьевич

к.т.н., доцент, Санкт-Петербургский политехнический университет Петра Великого
maleev_og@spbstu.ru

Аннотация. В статье предложен гибридный подход к построению диалоговой системы, сочетающий классические методы обработки естественного языка с современными нейросетевыми архитектурами. Предложена модульная архитектура, включающая классификатор намерений на основе рекуррентных сетей (GRU), модуль выделения именованных сущностей на базе трансформерной модели RuBERT и генеративный блок с использованием YandexGPT Lite. Для разрешения контекстной неоднозначности (эллипсисов) введён граф переходов, позволяющий определять вероятности тем с учётом истории диалога. Предложенная архитектура позволила достичь точности классификатора 0,92 при полноте 0,91. BLEU-оценки генерации лежат в диапазоне 0,19–0,33.

Ключевые слова: диалоговые системы, чат-бот, обработка естественного языка, нейронные сети, классификация намерений, распознавание именованных сущностей, трансформеры, эллипсисы, русский язык.

Введение

Современное развитие искусственного интеллекта привело к широкому внедрению диалоговых систем (чат-ботов) в бизнес-процессы, образование и повседневную жизнь [1]. Целевые (задаче-ориентированные) чат-боты позволяют автоматизировать ответ на типовые запросы пользователей. При этом ключевыми требованиями являются точное понимание намерения, извлечение необходимых параметров из запроса и естественность генерации ответа [2].

Русский язык представляет дополнительную сложность при проектировании системы, так как имеет богатую морфологию, свободный порядок слов и высокую контекстную зависимость многих высказываний [3]. В данной работе предлагается гибридный подход, объединяющий классические методы предобработки, нейросетевые классификаторы, трансформерные модели [4] для выделения сущностей и генеративные языковые модели. Предложен специальный блок обработки эллипсисов на основе графа переходов.

В настоящей работе рассматривается закрытая предметная область с фиксированным набором намерений, что характерно для большинства прикладных задач.

Архитектура разработанной диалоговой системы

Разработанная система состоит из пяти микросервисов: (1) хранилище датасетов, (2) классификатор намерений (NLU), (3) распознаватель сущностей (NER), (4) генератор ответов (на основе YandexGPT lite), (5) диалоговый менеджер, управляющий взаимодействием с пользователем.

Классификатор намерений

Вначале проводится предобработка поступившего в систему сообщения s , которая включает: удаление пунктуации, токенизацию, лемматизацию и удаление стоп-слов. В результате получаем последовательность лемм $t_1, \dots, t_n$. Каждая лемма преобразуется в вектор эмбединга $e_j \in \mathbb{R}^d$ с использованием сжатых FastText-

векторов [5]. Последовательность эмбедингов подаётся на слой GRU [6] нейронной сети, который вычисляет скрытые состояния $h_1, \dots, h_n$. Финальное состояние h_n пропускается через полносвязный слой с softmax для классификации интента по классам:

$$o = \text{softmax}(W_{out}h_n + b_{out}), \quad o_i = P(\text{intent} = i \mid s).$$

В демонстрационных целях было выбрано девять интенгов: приветствие, прощание, выражение благодарности, запрос помощи сотрудника, получение прогноза погоды, добавление встречи в календарь, удаление встречи, просмотр всех встреч на указанную дату, а также уточняющие вопросы (обработка эллипсисов).

Данные об интенгах хранятся в JSON-файле. Каждая запись содержит: уникальный идентификатор (tag); массив фраз-примеров (patterns) для обучения классификатора; список сущностей (entities), которые необходимо извлечь; поле ellipsis, указывающее на родительскую тему, если данный интент является эллипсисом.

Математическая модель графа переходов для эллипсисов

Эллипсис — это пропуск элемента высказывания, который восстанавливается из контекста. Например, после вопроса «Какая погода в Москве?» и ответа бота пользователь может спросить: «А в Самаре?». Без учёта контекста классификатор не сможет определить, относится ли этот вопрос к погоде или к другой теме.

Прямое добавление в датасет эллипсисов как отдельных интенгов приводит к коллизиям: фраза «А завтра?» может быть одинаково вероятна и для погоды, и для встреч. Чтобы разрешить эту неоднозначность, мы построили **граф переходов** контекста.

Введём множество интенгов $I = \{i_1, \dots, i_m\}$. Каждому интенгу i соответствует вектор признаков, но для текстовой обработки определим ориентированный взвешенный граф $G = (I, E, w)$, где $w : I \times I \rightarrow \mathbb{R}_{\geq 0}$ — функция веса ребра. Пусть $p_t(i)$ — вероятность, возвращаемая классификатором на шаге t для интенга i . Если предыдущий интент (на шаге $t - 1$) равен i^* , то пересчитанные вероятности для эллипсисов (подмножество $I_e \subset I$) задаются формулой:

$$\tilde{p}_t(i) = \begin{cases} p_t(i) \cdot w(i^*, i), & \text{если } i \in I_e, \\ p_t(i), & \text{иначе.} \end{cases}$$

Затем вероятности нормализуются (если необходимо). Веса w выбираются эмпирически: для релевантных переходов (погода → эллипсис погоды) $w > 1$, для нерелевантных (погода → эллипсис встреч) $w = 0$. Итоговый интент выбирается как $\arg\max_i \tilde{p}_t(i)$.

Если выбранный интент — эллипсис, заменить его родительским и подставить сохранённые параметры из предыдущего диалога.

Такой подход позволяет гибко настраивать связи между темами без переобучения классификатора.

Модуль NER на основе RuBERT

Для выделения сущностей использована модель RuBERT на основе русскоязычного варианта BERT[7]. Модель обучена распознавать 19 типов сущностей и показывает на русскоязычных данных общую полноту 79,4 %, причём для имён людей (PER) полнота достигает 95,7 %, для географических названий (LOC) — 82,6 %.

Генерация ответов

Генеративный модуль реализован на базе предобученной модели YandexGPT lite[8]. Обучение (дообучение) производилось на размеченном датасете, где для каждого интенга в поле «Сообщение» передавался либо оригинальный пользовательский запрос, либо ответ от внешнего API (например, текстовое описание погоды). Ожидаемый ответ генерировался моделью.

В ходе экспериментов YandexGPT lite не всегда точно воспроизводил конкретные параметры (даты, время, имена). Для интенгов, где требуется гарантированная достоверность (создание встречи, удаление встречи), ответ формируется из двух частей: основная фраза генерируется моделью, а точные данные вставляются с помощью шаблонизатора Jinja2. Это позволяет сохранить вариативность ответа, присущую генеративным моделям, и одновременно обеспечить точность.

Экспериментальная оценка

Для экспериментов сформирован датасет из 9 интенгов, каждый из которых содержит от 30 до 50 фраз-примеров. Общее количество размеченных предложений — 1000. Классификатор обучен с использованием оптимизатора Adam, скорость обучения 0.001, размер батча 32, 100 эпох. FastText-эмбединги сжаты до 196 МБ.

Метрики классификатора интенгов:

Метрика	Значение
Точность (Precision)	0,92
Полнота (Recall)	0,91

Для оценки качества генерации мы использовали метрику BLEU[9] для различных интенгов:

Интент	BLEU (среднее)
Погода	0,25
Помощь сотрудника	0,33
Список встреч	0,19
Добавление встречи	0,24
Уточнение параметров	0,29

Оценка обработки эллипсисов

Без графа переходов вероятность правильного распознавания эллипсиса «А завтра?» составляла 0,67 (с неоднозначностью между погодой и встречами). После применения графа с весами w (погода $\rightarrow$ элл. пого-

да) = 10 и $w(\text{встречи} \rightarrow \text{элл. погода}) = 0$ правильная вероятность возрастает до 0,95.

Проведено ручное тестирование корректного заполнения слотов и работу эллипсисов на 200 сценариях. Во всех случаях диалоговый менеджер успешно завершал интен- ты после сбора необходимых параметров.

Заключение

В статье представлена гибридный подход к построению диалоговой системы. Предложено внедрение графа переходов для обработки эллипсисов. Эксперименты подтвердили эффективность подхода: точность классификации намерений 0,92, полнота 0,91; BLEU-оценки генерации ответов в диапазоне 0,19–0,33.

ЛИТЕРАТУРА

1. Веденева Е.Д., Малеев О.Г. Разработка алгоритма для автоматической оценки качества работы специалистов клиентской службы на основе больших языковых моделей // Информационные технологии. Проблемы и решения. 2025. Т. 31, № 2. С. 5–11. EDN XCB1HB.
2. Young S. и др. POMDP-Based Statistical Spoken Dialog Systems: A Review // Proc. IEEE. IEEE, 2013. Т. 101, № 5. С. 1160–1179.
3. Romanov V., Khusainova A. Evaluation of Morphological Embeddings for the Russian Language // Proceedings of the 2019 3rd International Conference on Natural Language Processing and Information Retrieval. New York, NY, USA: ACM, 2019. С. 144–148.
4. Vaswani A. и др. Attention is all you need // Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2017. С. 6000–6010.
5. Bojanowski P. и др. Enriching Word Vectors with Subword Information // Trans. Assoc. Comput. Linguist. MIT Press, 2017. Т. 5. С. 135–146.
6. Cho K. и др. Learning phrase representations using RNN encoder-decoder for statistical machine translation // EMNLP 2014 — 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference. 2014. С. 1724–1734.
7. Devlin J. и др. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // NAACL HLT 2019 — 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. — Proc. Conf. Association for Computational Linguistics (ACL), 2018. Т. 1. С. 4171–4186.
8. Moshninov A., Komarov M. Artificial Intelligence in Russia: Evaluating Yandex GPT-5 and GigaChat 2.0 MAX in Global Context // ICAIMT. Abu Dhabi, 2025. С. 38–44.
9. Papineni K. и др. BLEU: a method for automatic evaluation of machine translation // Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. USA: Association for Computational Linguistics, 2002. С. 311–318.

© Малеев Олег Геннадьевич (maleev_og@spbstu.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

СРАВНИТЕЛЬНЫЙ АНАЛИЗ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ СУПЕРРАЗРЕШЕНИЯ ИЗОБРАЖЕНИЙ

COMPARATIVE ANALYSIS OF NEURAL NETWORK MODELS OF IMAGE SUPER RESOLUTION

D. Mirsagatov

Summary. This paper compares single-image super-resolution models from the ESRGAN family: ESRGAN, A-ESRGAN, and Real-ESRGAN. The experiment is conducted on a small subset of the DIV2K dataset: HR images are degraded to LR using bicubic downscaling ($\times 4$) and then upsampled back by each model. Quality is evaluated using PSNR, SSIM, and LPIPS metrics along with visual comparison. On the selected test set, Real-ESRGAN achieves the best mean scores (PSNR = 22.764 dB; SSIM = 0.6452; LPIPS = 0.2327), followed by A-ESRGAN and ESRGAN. Limitations related to the small sample size and synthetic degradation are discussed.

Keywords: single-image super-resolution, GAN, ESRGAN, Real-ESRGAN, DIV2K, PSNR, LPIPS.

Мирсагатов Даврон Фуркатович

ФБГОУ ВО Поволжский государственный университет
телекоммуникаций и информатики, Самара
davron.mirsagatovv@mail.ru

Аннотация. В работе выполнено сравнение нейросетевых моделей однокадрового суперразрешения изображений семейства ESRGAN: ESRGAN, A-ESRGAN и Real-ESRGAN. Эксперимент проведён на подвыборке набора DIV2K: исходные HR-изображения деградировались до LR с помощью bicubic downscale $\times 4$, после чего восстанавливались до масштаба $\times 4$ каждой моделью. Качество оценивалось метриками PSNR, SSIM и LPIPS, а также визуальным сопоставлением результатов. По средним значениям метрик на выбранном наборе наилучший результат показал Real-ESRGAN (PSNR = 22,764 дБ; SSIM = 0,6452; LPIPS = 0,2327), далее A-ESRGAN и ESRGAN. Отмечены ограничения исследования, связанные с малым объёмом выборки и использованием синтетической деградации.

Ключевые слова: суперразрешение, однокадровое суперразрешение, генеративно-состязательные сети, ESRGAN, Real-ESRGAN, PSNR, LPIPS.

Введение

Задача суперразрешения изображения (SISR) заключается в восстановлении изображения высокого разрешения по его версии с низким разрешением. Практическая значимость данной задачи обусловлена широким спектром применений: улучшение потребительских фото и видео, повышение информативности изображений в медицине и промышленной диагностике, восстановление архивных материалов, а также подготовка контента для печати и цифровых витрин. При этом увеличение разрешения классическими методами интерполяции (например, бикубической) не восстанавливает утраченные детали и часто приводит к размытию границ и ухудшению визуальной чёткости [10, 6].

В научной литературе представлены как классические подходы к суперразрешению [10, 6], так и нейросетевые методы на базе свёрточных сетей и генеративно-состязательных моделей [5, 1]. В данной работе рассматриваются модели семейства ESRGAN (A-ESRGAN и Real-ESRGAN) и выполняется их сравнительное исследование.

Современные методы обработки изображений на основе глубокого обучения позволяют не только повышать разрешение, но и восстанавливать мелкие детали за счёт обучения на больших наборах данных. Особое

место занимают генеративно-состязательные подходы (GAN), ориентируемые на перцептивное качество и визуальную правдоподобность текстуры. Однако такие методы могут создавать артефакты и «дорисованные» детали, поэтому для корректного сравнения требуется сочетание количественных метрик и визуальной оценки.

Цель работы — выполнить сравнительный анализ трёх моделей семейства ESRGAN (ESRGAN, A-ESRGAN и Real-ESRGAN) как эволюционной линии GAN-подходов к однокадровому суперразрешению (SISR) и определить различия их поведения на единой экспериментальной базе по метрикам PSNR, SSIM, LPIPS и по визуальному качеству.

Далее в статье кратко рассматриваются ключевые архитектурные идеи выбранных моделей, описывается экспериментальный протокол (HR-эталон, bicubic деградация $\times 4$), приводятся количественные метрики и визуальные сравнения, после чего формулируются выводы и ограничения исследования.

Обзор методов

1. ESRGAN (Enhanced Super-Resolution GAN)

ESRGAN относится к классу генеративно-состязательных методов суперразрешения. Основная идея

заключается в использовании генератора, который восстанавливает изображение высокого разрешения, и дискриминатора, который отличает восстановленные изображения от реальных. За счёт состязательного обучения и перцептивной функции потерь ESRGAN способен восстанавливать визуально правдоподобные текстуры и детали, превосходя интерполяционные методы по субъективному восприятию [1].

Ключевая особенность ESRGAN — использование остаточных блоков Residual-in-Residual Dense Blocks (RRDB) в генераторе и переход к перцептивно-ориентированному обучению (перцептивная и состязательная составляющие), что помогает получать более детализированные и визуально правдоподобные результаты. [1] Обучение ESRGAN обычно оптимизируется по комбинации нескольких функций потерь:

- пиксельная (content) потеря для сохранения общей структуры;
- перцептивная потеря (например, по признакам сверточной сети), коррелирующая с человеческим восприятием;
- состязательная (adversarial) потеря для повышения фотореалистичности и «резкости» текстур.

Сильная сторона ESRGAN — высокое перцептивное качество. Ограничение — возможность появления артефактов и «галлюцинаций» деталей при несоответствии распределения деградаций обучающим данным.

2. A-ESRGAN (Adaptive ESRGAN)

В рамках настоящей работы A-ESRGAN рассматривается не как отдельная каноническая архитектура, а как предобученный вариант генератора типа RRDBNet x4, унаследованный от ESRGAN-семейства. Отличие от базового ESRGAN в эксперименте заключается не в структуре сети, а в используемом наборе весов (чекпойнте) и, как следствие, в ином компромиссе между подавлением шумов/артефактов и сохранением мелких деталей. Используемый чекпойнт: A_ESRGAN_Single.pth [12].

3. Real-ESRGAN

Real-ESRGAN — развитие идей ESRGAN, ориентированное на сценарии «реального мира», когда входные изображения имеют сложные деградации: шумы, компрессию, размытия, артефакты масштабирования и смешанные искажения. В задаче однокадрового суперразрешения модель получает на вход изображение низкого разрешения (LR) и восстанавливает по нему изображение более высокого разрешения (SR/HR). Для обучения и объективного сравнения LR часто формируют искусственно из эталонного HR путём понижения разрешения (например, bicubic downsampling). Real-ESRGAN отличается тем, что при обучении использует более ре-

алистичную модель деградаций (шум, компрессия, размытие) вместо «простого» downsampling, что повышает устойчивость на данных реального мира.

Согласно работе X. Wang Real-ESRGAN ориентирован на более устойчивое восстановление в условиях «реальных» деградаций (шум, компрессия, размытие), что снижает риск выраженных артефактов по сравнению с агрессивными GAN-подходами. При этом в отдельных случаях перцептивная «резкость» может быть ниже, зато повышается предсказуемость результата [2].

Метрики оценки качества

Для количественного сравнения результатов суперразрешения в работе применяются метрики PSNR, SSIM и LPIPS. Все три метрики относятся к оценкам качества при наличии эталонного изображения (full-reference). Поэтому эксперимент строится по схеме: изображение высокого разрешения (HR) → деградация до низкого разрешения (LR) → восстановление (SR) → сравнение SR с HR.

1. **PSNR (Peak Signal-to-Noise Ratio)** — метрика, основанная на среднеквадратичной ошибке (MSE) между восстановленным изображением и эталоном. PSNR измеряется в децибелах; большее значение обычно соответствует меньшей пиксельной ошибке. Метрика хорошо отражает точность на уровне пикселей, но может слабо коррелировать с субъективным качеством для GAN-методов, которые способны «дорисовывать» детали, изменяя пиксельные значения.
2. **SSIM (Structural Similarity Index)** оценивает структурное сходство двух изображений, учитывая локальные статистики яркости, контраста и структуры. Значение SSIM лежит в диапазоне от 0 до 1 (в типичной нормировке), где ближе к 1 означает большее структурное сходство с эталоном. SSIM лучше PSNR отражает сохранение структуры, но также не всегда полностью описывает перцептивные свойства результата.
3. **LPIPS (Learned Perceptual Image Patch Similarity)** — перцептивная метрика, основанная на расстоянии между признаками, извлекаемыми нейросетевой моделью. Она лучше коррелирует с человеческим восприятием различий, особенно в задачах улучшения визуального качества. Для LPIPS меньшее значение означает большее перцептивное сходство восстановленного изображения с эталоном.

В работе метрики используются совместно: PSNR и SSIM отражают точность и структурную близость, а LPIPS — перцептивное сходство. Дополнительно выполняется экспертная визуальная оценка человеком

Таблица 1.
Набор тестовых изображений

Изображение	Тип сцены	Размер HR-изображения после предобработки (высота×ширина), пикс.	Размер LR-изображения(×4) (высота×ширина), пикс.
0001.png	макросьёмка / объект (морская звезда)	704×1024	176×256
0002.png	сцена / пейзаж (пальмы, город/пляж)	680×1024	170×256
0003.png	пейзаж (водоём/берег)	672×1024	168×256
0004.png	макросьёмка / объект (бабочка)	1024×680	256×170
0005.png	пейзаж (горы/долина)	680×1024	170×256

(наблюдателем), так как для GAN-подходов возможны случаи, когда формальные метрики и субъективное восприятие расходятся.

Эксперимент

Для сравнения методов использовался набор из 5 HR-изображений датасета DIV2K [12]: 0001.png, 0002.png, 0003.png, 0004.png и 0005.png (см. табл. 1). Изображения представлены в формате PNG и не подвергались дополнительному сжатию с потерями. Для повышения воспроизводимости и снижения времени вычислений исходные изображения масштабировались так, чтобы максимальная сторона не превышала 1024 пикселя; далее выполнялось центрированное кадрирование до размеров, кратных 4.

LR-изображения формировались методом бикубической интерполяции (bicubic downscale) с коэффициентом 4, т.е. $LR = HR/4$. Полученные LR-изображения подавались на вход каждой модели (ESRGAN, A-ESRGAN и Real-ESRGAN), после чего модели формировали SR-результат масштаба ×4. Метрики PSNR (дБ), SSIM и LPIPS вычислялись между SR и соответствующим эталоном HR.

HR (эталон) → деградация (bicubic downscale ×4) → LR (вход)
LR → {ESRGAN, A-ESRGAN, Real-ESRGAN} → SR (×4) → метрики (SR vs HR): PSNR, SSIM, LPIPS

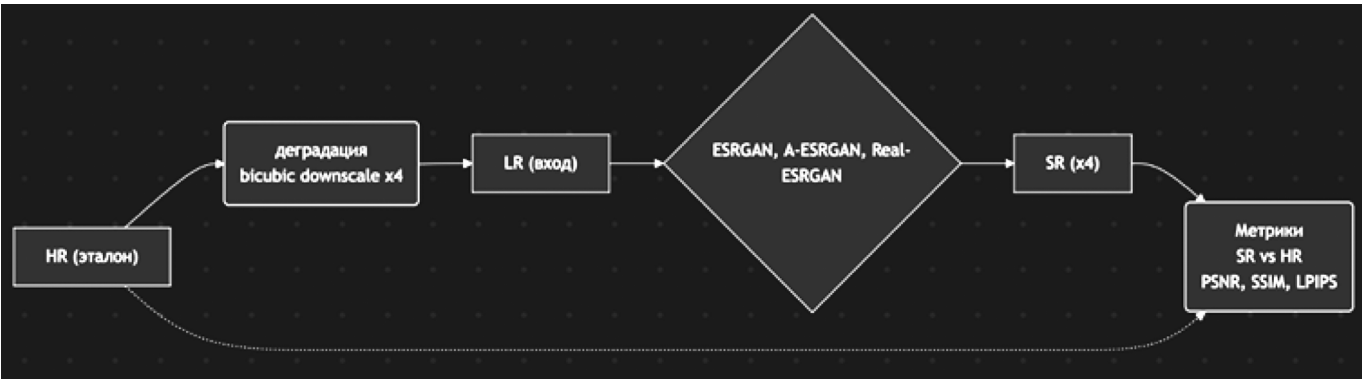


Рис. 1. Схема экспериментального протокола (HR → LR → SR → метрики).

Таблица 2.

Результаты эксперимента по каждому изображению (PSNR, SSIM, LPIPS)

Модель	ESRGAN (дБ ↑)	A-ESRGAN (дБ ↑)	Real-ESRGAN (дБ ↑)	ESRGAN (↑)	A-ESRGAN (↑)	Real-ESRGAN (↑)	ESRGAN (↓)	A-ESRGAN (↓)	Real-ESRGAN (↓)
Метрика	PSNR	PSNR	PSNR	SSIM	SSIM	SSIM	LPIPS	LPIPS	LPIPS
0001.png	13.150	20.400	22.681	0.1671	0.4441	0.5334	0.8522	0.3642	0.3582
0002.png	12.436	19.311	19.938	0.1889	0.5322	0.5694	0.7201	0.2811	0.2469
0003.png	18.804	24.835	25.253	0.4961	0.8048	0.8283	0.5024	0.1694	0.1483
0004.png	18.259	24.738	26.540	0.4588	0.8085	0.8466	0.4725	0.1634	0.1292
0005.png	12.424	19.503	19.410	0.1383	0.4291	0.4485	0.8604	0.2883	0.2810
Ср. ариф.	15.015	21.758	22.764	0.2898	0.6038	0.6452	0.6815	0.2533	0.2327

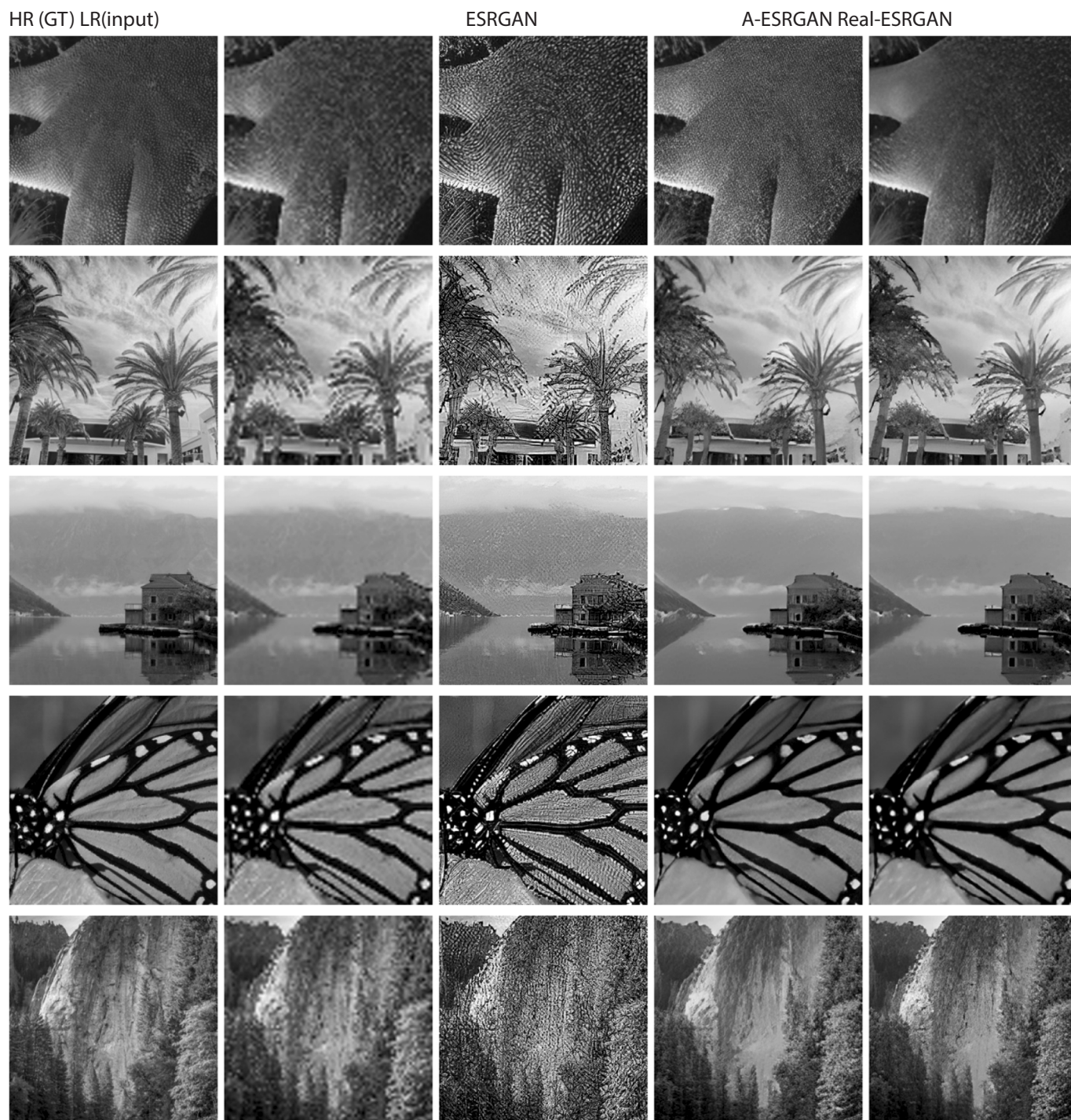
Примечание: в работе использовались предобученные веса (чекпойнты) для масштаба $\times 4$: ESRGAN_SRx4_DF2KOST_official для ESRGAN, A_ESRGAN_Single.pth для A-ESRGAN и RealESRGAN_x4plus.pth для Real-ESRGAN (архитектура RRDBNet).

Анализ результатов

Анализ результатов выполнялся по трём количественным метрикам: PSNR (дБ; больше — лучше), SSIM (больше — лучше) и LPIPS (меньше — лучше). В табл. 2–4

представлены значения PSNR, SSIM и LPIPS для каждого изображения, что позволяет оценить устойчивость методов на разных сценах. Для сводного сравнения рассчитаны средние значения по набору (табл. 5), однако из-за малого объёма выборки (5 изображений) их следует рассматривать как ориентир и дополнять выводы визуальным сопоставлением результатов.

По средним значениям (табл. 5) на выбранной подвыборке DIV2K наилучшие показатели демонстрирует Real-ESRGAN (PSNR = 22.764 дБ; SSIM = 0.645; LPIPS =



0.233). A-ESRGAN занимает промежуточное положение (PSNR = 21.758 дБ; SSIM = 0.604; LPIPS = 0.253), а базовый ESRGAN уступает по всем метрикам (PSNR = 15.015 дБ; SSIM = 0.290; LPIPS = 0.682). Полученный порядок ожидаем и согласуется с тем, что Real-ESRGAN является более поздней модификацией семейства и ориентирован на устойчивость к более реалистичным деградациям. По отдельным изображениям разброс метрик заметен: на сценах с мелкой текстурой и контрастными границами преимущество Real-ESRGAN выражено сильнее, тогда как базовый ESRGAN чаще даёт артефакты и хуже сохраняет структуру.

Заключение

В работе выполнено сравнение моделей семейства ESRGAN (ESRGAN, A-ESRGAN и Real-ESRGAN) в воспроизводимом протоколе суперразрешения $\times 4$ на подвыборке изображений из DIV2K. Количественная оценка

выполнялась по метрикам PSNR, SSIM и LPIPS, а также проводилось визуальное сопоставление результатов. Полученные значения показывают, что в данном протоколе и на выбранном наборе Real-ESRGAN обеспечивает лучшие средние показатели, A-ESRGAN даёт компромисс между сохранением структуры и подавлением артефактов, а базовый ESRGAN чувствителен к выбранным настройкам и типу деградации.

Ограничения исследования связаны с небольшим объёмом тестового набора (5 изображений) и использованием упрощённой схемы деградации bicubic downscale, которая не полностью отражает реальные искажения. В дальнейшем целесообразно расширить набор изображений и протестировать дополнительные типы деградаций (например, JPEG-компрессию, шум, размытие), а также рассмотреть влияние выбора чекпойнта и настроек инференса на метрики и визуальный результат.

ЛИТЕРАТУРА

1. Wang X., Yu K., Wu S. и др. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. arXiv:1809.00219, 2018.
2. Wang X., Xie L., Dong C., Shan Y. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. arXiv:2107.10833, 2021.
3. Zhang R., Isola P., Efros A. A., Shechtman E., Wang O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric (LPIPS). arXiv:1801.03924, 2018.
4. Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Trans. Image Processing, 2004.
5. Ledig C., Theis L., Huszar F. и др. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network (SRGAN). arXiv:1609.04802, 2017.
6. Dong C., Loy C.C., He K., Tang X. Image Super-Resolution Using Deep Convolutional Networks (SRCNN). arXiv:1501.00092, 2015.
7. Lim B., Son S., Kim H., Nah S., Mu Lee K. Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR). arXiv:1707.02921, 2017.
8. Yang J., Wright J., Huang T. S., Ma Y. Image Super-Resolution via Sparse Representation. IEEE Trans. Image Processing, 2010.
9. Freeman W.T., Jones T.R., Pasztor E.C. Example-Based Super-Resolution. IEEE Computer Graphics and Applications, 2002.
10. Keys R.G. Cubic Convolution Interpolation for Digital Image Processing. IEEE Trans. Acoustics, Speech, and Signal Processing, 1981.
11. Tsai R.Y., Huang T.S. Multiframe Image Restoration and Registration. Advances in Computer Vision and Image Processing, 1984.
12. Wei Z., Huang Y., Chen Y., Zheng C., Gao J. A-ESRGAN: Training Real-World Blind Super-Resolution with Attention U-Net Discriminators. arXiv:2112.10046, 2021.

МОДЕЛЬ УСТОЙЧИВОГО МАРКИРОВАНИЯ СЕТЕВЫХ ПОТОКОВ ДЛЯ ПОВЫШЕНИЯ ЗАЩИЩЁННОСТИ ЧАСТИЧНО КОНТРОЛИРУЕМЫХ СЕТЕЙ

Молькова Лолита Юрьевна

ФГАОУ ВО «Санкт-Петербургский политехнический
университет Петра Великого»

lolita-molkova@mail.ru

A MODEL FOR ROBUST NETWORK FLOW MARKING FOR ENHANCING THE SECURITY OF PARTIALLY CONTROLLED NETWORKS

L. Molkova

Summary. This paper studies an interval-based network flow marking method aimed at improving marker detection robustness in partially controlled networks, where the true source of a connection may be hidden behind a chain of intermediate nodes. The method follows the patchwork principle: each information bit is encoded through the relation between packet counts in adjacent time intervals. Two marker sequences are compared: a pseudorandom bipolar sequence and a limit bipolar expanding sequence. Four attack scenarios are considered in the evaluation: flow mixing, packet loss, jitter, and packet repackaging. The results show that the use of the limit bipolar expanding sequence reduces the probability of marker miss detection in most scenarios while maintaining a comparable level of false alarms. The most pronounced advantage is observed under jitter and flow mixing.

Keywords: interval-based marking, patchwork method, network flow watermark, packet-count watermarking, robustness to attacks, partially controlled network, stepping stones.

Аннотация. В работе рассмотрен интервальный метод маркирования сетевых потоков, предназначенный для повышения устойчивости обнаружения в частично контролируемых сетях, где источник соединения может маскироваться цепочкой промежуточных узлов. Метод реализует принцип «лоскута»: каждый информационный бит кодируется соотношением количества пакетов в соседних временных интервалах. В качестве маркерной последовательности сопоставлены два варианта — псевдослучайная биполярная последовательность и предельная биполярная расширяющаяся последовательность. Для оценки устойчивости рассмотрены четыре модели воздействия: смешивание потоков, потери пакетов, временной джиттер и переупаковка пакетов. Показано, что использование предельной биполярной расширяющейся последовательности уменьшает вероятность пропуска маркера в большинстве сценариев при сопоставимом уровне ложных срабатываний. Наиболее выраженный выигрыш наблюдается при джиттере и смешивании потоков.

Ключевые слова: интервальное маркирование, метод лоскута, сетевой водяной знак, маркирование сетевого потока, устойчивость к искажениям, частично контролируемая сеть, промежуточные узлы.

Введение

В частично контролируемых сетях анализатор трафика обычно видит только отдельные фрагменты маршрута. Это усложняет идентификацию нарушителя, поскольку реальный источник соединения может быть скрыт за цепочкой промежуточных узлов, прокси-серверов или stepping-stone-переходов. При таком сценарии пассивной корреляции трафика часто недостаточно: временные характеристики соединения искажаются, а журналы событий оказываются неполными или трудно сопоставимыми.

Одним из активных подходов к решению этой задачи является маркирование сетевых потоков — контролируемое внесение малозаметной сигнатуры в структуру трафика с последующим её обнаружением на другом участке сети [1, 2, 19]. В последние годы исследовались как схемы активного маркирования для задач деано-

нимизации (deanonymization) и трассировки источника трафика (traceback) [3–10], так и методы повышения устойчивости к помехам, синхронизационным смещениям и атакующим преобразованиям [12–18].

Для интервальных методов особенно важны три требования: скрытность маркера, устойчивость к сетевым искажениям и корректная синхронизация при детектировании. В работах последних лет это реализуется по-разному: через динамические интервалы, неоднородные временные каналы, смешанные признаки последовательности и времени, а также адаптивную настройку параметров детектора [7–10, 13, 14, 17, 18]. Вместе с тем в частично контролируемых сетях остаётся актуальной задача построения схемы, которая сохраняет различимость маркера при сравнительно простом алгоритме внедрения и восстановления.

В настоящей работе рассматривается интервальный метод потокового маркирования, основанный на прин-

ципе лоскута и ПБРП. Для обеспечения более надёжной идентификации встраиваемого идентификатора в работе применяется предельная биполярная расширяющаяся последовательность (ПБРП), детальный анализ свойств которой представлен в исследовании [11]. Использование данного подхода позволяет достичь компромисса между простотой реализации процедуры встраивания и улучшенными корреляционными характеристиками применяемых сигнатур. Полученные в ходе работы результаты также сопоставляются с выводами современных обзорных и прикладных исследований в области маркирования сетевого трафика [15–19].

Методы

1. Анализ современных подходов

Среди современных решений можно выделить несколько групп методов. Первая группа использует интервально-временное преобразование трафика: водяной знак формируется за счёт изменения распределения пакетов по интервалам или за счёт модуляции межпакетных задержек [1–4, 8]. Вторая группа делает акцент на устойчивую синхронизацию и динамический выбор параметров внедрения [7, 9, 12, 18]. Третья группа комбинирует временные признаки с порядком пакетов, что особенно заметно в схемах HSTW и IP-Peeling [13, 14].

Для рассматриваемой задачи существенны не столько максимальная пропускная способность маркера или его скрытность в общем случае, сколько способность детектора сохранять различимость при локальных искажениях потока. Поэтому в данной статье в качестве базовой модели выбран интервальный метод, основанный на методе лоскута и ПБРП. Такая постановка удобна для формального описания, а также для моделирования атак смешивания, потерь, джиттера и переупаковки.

Дополнительно вместо традиционной псевдослучайной последовательности в качестве опорного сигнала рассматривается ПБРП. Для лоскутного принципа это особенно важно, поскольку качество обнаружения определяется не только уровнем искажения потока, но и корреляционными свойствами самой маркерной последовательности [11, 15, 16].

2. Математическая модель интервального маркирования

Вектор сетевых пакетов определяется формулой [15]:

$$\mathbf{p} = (p_0, p_1, \dots, p_N), \quad (1)$$

где p_i — число пакетов в i -м интервале, а N — число интервалов.

Для имитационного моделирования интенсивность потока задаётся пуассоновской случайной величиной.

$$p_i = \text{mod}(\xi_i, M), \quad (2)$$

где $\xi_i \sim \text{Poisson}(\lambda)$ — случайная величина, подчиняющаяся **пуассоновскому распределению** с параметром интенсивности λ ; M — максимально допустимое число пакетов в интервале.

3. Формирование маркерной последовательности

Маркер задаётся биполярной последовательностью длины N .

$$\mathbf{m} = (m_1, m_2, \dots, m_N) = \alpha \otimes (1, -1), \quad (3)$$

где $\otimes$ — операция кронекерова произведения [15], $\alpha = \left(\alpha_0, \alpha_1, \dots, \alpha_{\frac{N}{2}} \right)$ — биполярный вектор, элементы которого принимают значения $\{-1, 1\}$ [15].

В работе сравниваются два варианта опорной последовательности. В первом случае α формируется как псевдослучайная биполярная последовательность. Во втором используется ПБРП [11], которая строится по кронекеру произведению двух биполярных векторов:

$$\alpha = \beta \otimes \gamma, \quad (4)$$

В вычислительных экспериментах использовались последовательности $\beta = (1, 1, 1, -1, -1, 1, -1)$ и $\gamma = (-1, -1, 1, -1)$, поэтому длина результирующей последовательности составляла $N_\alpha = N_\beta \cdot N_\gamma = 28$ [15]. Такой выбор позволяет сравнивать ПБРП и псевдослучайный маркер при одинаковой длине опорного сигнала.

4. Внедрение маркера по принципу лоскута

Формирование маркированного вектора пакетов выполняется по формуле (1) [15].

5. Восстановление маркера и правило решения

После прохождения сетевой среды выполняется восстановление маркерной последовательности по результирующему потоку:

$$\mathbf{p}_{\text{marker}} = (p'_0, p'_1, \dots, p'_N), \quad (5)$$

где p'_i — изменённые значения пакетов после внедрения маркера.

Процесс построения $\mathbf{p}_{\text{marker}}$ осуществляется таким же образом, что и в формуле (3–4).

6. Модель восстановления маркера

После прохождения сетевой среды и добавления шумового потока p_2 результирующий поток описывается выражением:

$$p_{out} = p_{marker} + p_2, \quad (6)$$

где $p_2 = (p_0^{(2)}, p_1^{(2)}, \dots, p_N^{(2)})$ — вектор случайных пакетов, сгенерированных также по пуассоновскому распределению с параметром λ . В данной модели вектор p_2 можно рассматривать, как шумовой вектор.

Восстановление маркера выполняется путём анализа изменения временной структуры между соседними парами пакетов результирующего потока p_{out} . Для каждой пары вычисляется разность между вторым и первым элементом одной и той же пары результирующего потока:

$$\Delta_i = p_{out,2i} - p_{out,2i-1}, \quad (7)$$

на основании которой формируется восстановленный маркер:

$$\tilde{m}_i = \begin{cases} 1, \Delta_i < 0 \\ -1, \Delta_i \geq 0 \end{cases}, \quad (8)$$

Такой способ анализа учитывает взаимное влияние соседних пар пакетов, что повышает точность детектирования маркера при наличии аддитивных шумов и временных искажений в сетевом канале.

Восстановленная последовательность $\tilde{m}_n$ сопоставляется с оригинальной m_n с помощью взаимной корреляции по сдвигу [15]:

$$R(\tilde{m}, m, k) = \sum_n (\tilde{m}_n \cdot m_{n+k}), \quad (9)$$

Здесь суммирование ведётся по допустимым индексам, а k задаёт проверяемый сдвиг маркера. Для принятия решения используется отношение максимального по модулю значения корреляции к минимальному [15]:

$$A = \left| \frac{\max_{k \in \{-N, -N+1, \dots, N\}} R(\tilde{m}, m, k)}{\min_{k \in \{-N, -N+1, \dots, N\}} R(\tilde{m}, m, k)} \right|, \quad (10)$$

где $|\cdot|$ — оператор взятия абсолютного значения; A — оценка взаимной корреляции между восстановленным маркером и оригинальным маркером; R — значение корреляции между восстановленным маркером и оригинальным маркером.

При выполнении условия $A \geq \tau$ делается вывод о присутствии маркера в исследуемом потоке данных. Вели-

чина порогового значения τ определяется исходя из необходимости обеспечения баланса между вероятностью пропуска имеющегося маркера и вероятностью ложного обнаружения.

7. Модели деструктивных воздействий на маркированный трафик

В рамках настоящей работы рассматривается четыре типа угроз, характерных для сред с ограниченным уровнем контроля. Данные модели учитывают как модификацию интенсивности трафика, так и искажение его временной структуры.

7.1. Атака типа «смешивание»

Атака типа «смешивание» моделирует ситуацию, при которой анализируемый поток формируется как линейная комбинация целевого маркированного трафика и постороннего фоновой трафика, не содержащего маркера. Подобный сценарий возникает при объединении нескольких независимых потоков на промежуточном сетевом узле или при прохождении трафика через общий канал связи.

$$y_i = (1 - a)p'_i + az_i, \quad (11)$$

где z_i — независимый фоновый поток $z_i = p_i^{(2)}$; $a \in [0, 1]$ — коэффициент смешивания.

Чем ближе a к единице, тем меньше вклад маркированного потока и тем труднее его последующее обнаружение.

Чем ближе a к единице, тем меньше вклад маркированного потока и тем труднее его последующее обнаружение.

7.2. Потери пакетов

Потери пакетов моделируются как независимое удаление пакетов из каждого интервала с вероятностью p_{loss} . Тогда число пакетов, дошедших до наблюдателя, описывается биномиальной случайной величиной:

$$p_{loss} \in [0, 1], \quad (12)$$

Пусть p'_i число пакетов в i -м временном интервале маркированного потока. Тогда число пакетов, успешно дошедших до наблюдателя, является случайной величиной y_i , которая подчиняется **биномиальному распределению**:

$$y_i \sim \text{Binomial}(p'_i, 1 - p_{loss}), \quad (13)$$

Биномиальное распределение $Binomial(n, p)$ описывает число успешных исходов в n независимых испытаниях Бернулли, где вероятность успеха в каждом испытании равна p . В предлагаемой модели:

- $n = p'_i$ — количество переданных пакетов,
- $p = 1 - p_{loss}$ — вероятность успешной доставки пакета.

Следовательно, атака приводит к ослаблению интенсивности потока и уменьшению разности между соседними интервалами пары.

7.3. Временной джиттер

В данной работе джиттер моделируется как случайное возмущение числа пакетов в интервале, отражающее влияние временных флуктуаций канала на распределение пакетов по временным окнам.

Пусть добавка η_i — центрированный шум:

$$y_i = p'_i + \eta_i, i = 0, \dots, N, \quad (14)$$

где $\eta_i = \xi_i + \lambda_j$ при этом λ_j — интенсивность джиттера и $\xi_i \sim Poisson(\lambda_j) - \lambda_j$.

7.4. Атака типа «Перераспределение или Переупаковка пакетов»

Воздействие, обозначаемое как переупаковка, описывает процесс перераспределения части пакетов между смежными временными интервалами. Данное явление может быть следствием таких операций, как агрегация, фрагментация или изменение формата пакетов на сетевых устройствах. Математически данное воздействие описывается следующим соотношением:

$$y_i = \beta p'_i + (1 - \beta) p'_{i-1} \beta \in [0, 1] \quad (15)$$

В приведённой формуле параметр β характеризует интенсивность перераспределения пакетов между соседними интервалами и определяет степень размывания бинарной структуры встроенного маркера. Уменьшение значения β ведёт к более существенному влиянию данных из предыдущего интервала, что усиливает искажение структуры маркера.

8. Условия проведения экспериментальных исследований

Для каждого из рассматриваемых сценариев было выполнено 2000 независимых имитационных экспериментов. Базовая интенсивность пуассоновского потока была установлена на уровне $\lambda = 500$. Параметры моделируемых атак имели следующие значения: коэффициент перераспределения для атаки переупаковки составлял 0.8; интенсивность флуктуаций при джиттере была равна 1; доля примешиваемого постороннего трафика при смешивании устанавливалась на уровне 0.5; вероятность потери пакета соответствовала значению 0.05 (5 %). Для обеспечения корректности сравнения оба исследуемых маркера имели одинаковую длину, равную 28 битам, что позволило оценивать их устойчивость при равном объёме встраиваемой служебной информации.

Результаты моделирования

На рисунках 1–8 представлены зависимости вероятности пропуска маркера и вероятности ложного обнаружения от величины порога τ для четырёх типов рассматриваемых воздействий. На всех графиках проводится сравнение эффективности использования псевдослучайной биполярной последовательности и ПБРП. Увеличение порогового значения закономерно приводит к снижению вероятности ложного обнаружения, однако одновременно возрастает риск пропуска имеющегося маркера. Выбор конкретного значения τ должен определяться требуемым балансом между указанными показателями.

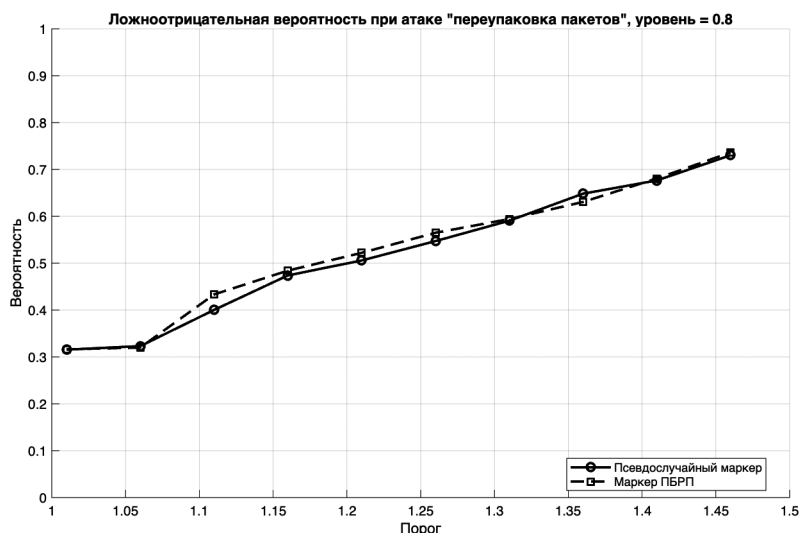


Рис. 1. Зависимость вероятности пропуска маркера от порога τ при атаке переупаковки пакетов

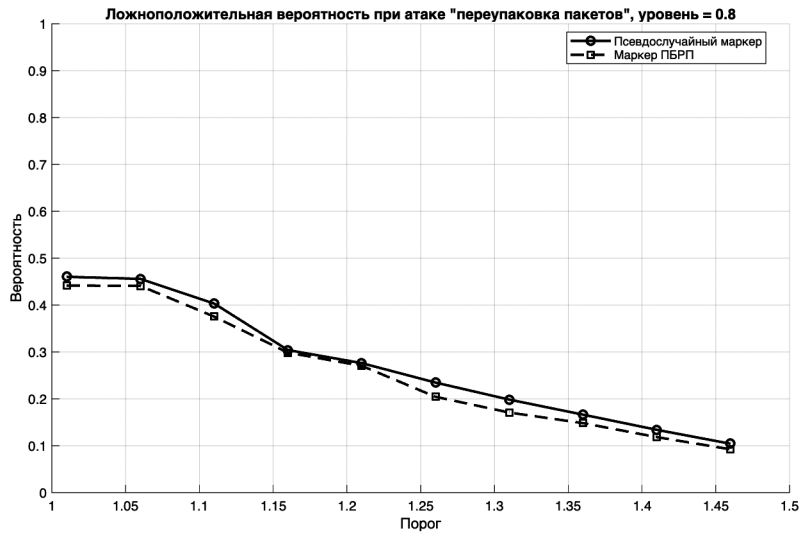


Рис. 2. Зависимость вероятности ложного срабатывания от порога τ при атаке переупаковки пакетов

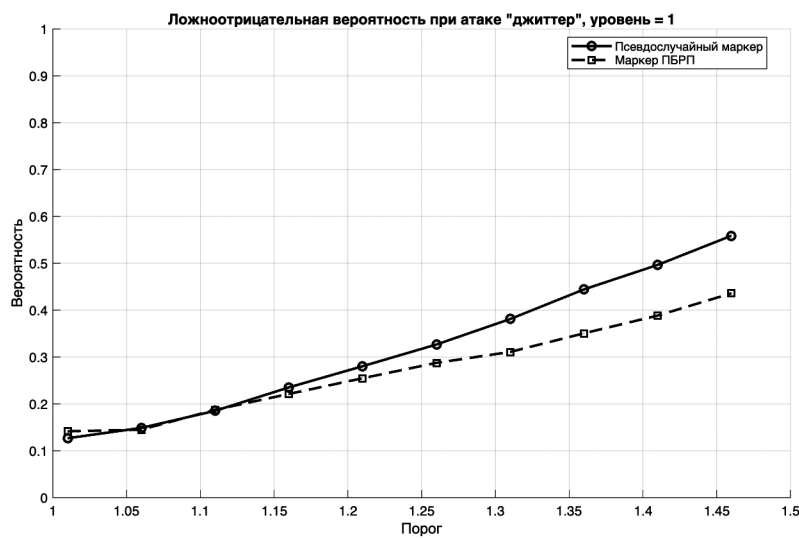


Рис. 3. Зависимость вероятности пропуска маркера от порога τ при атаке временного джиттера

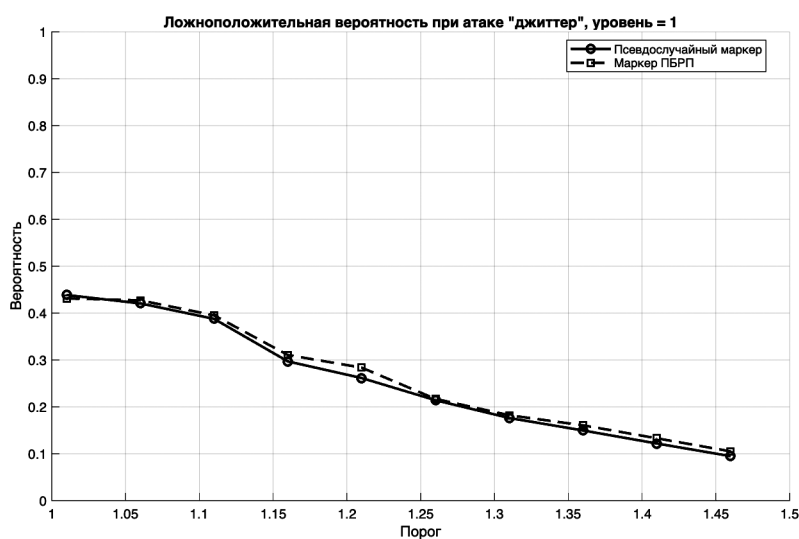
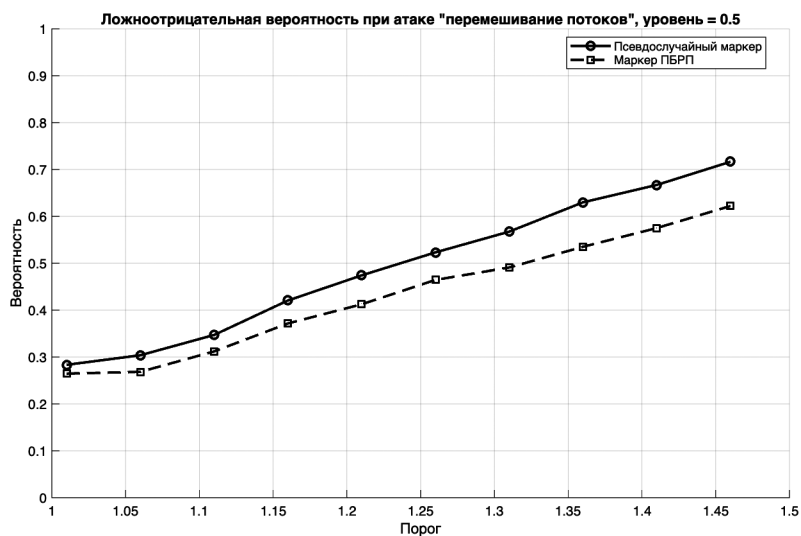
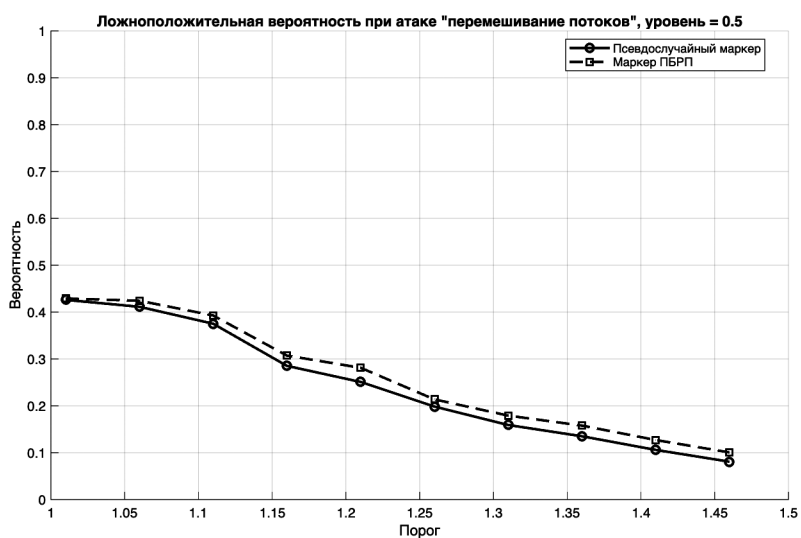
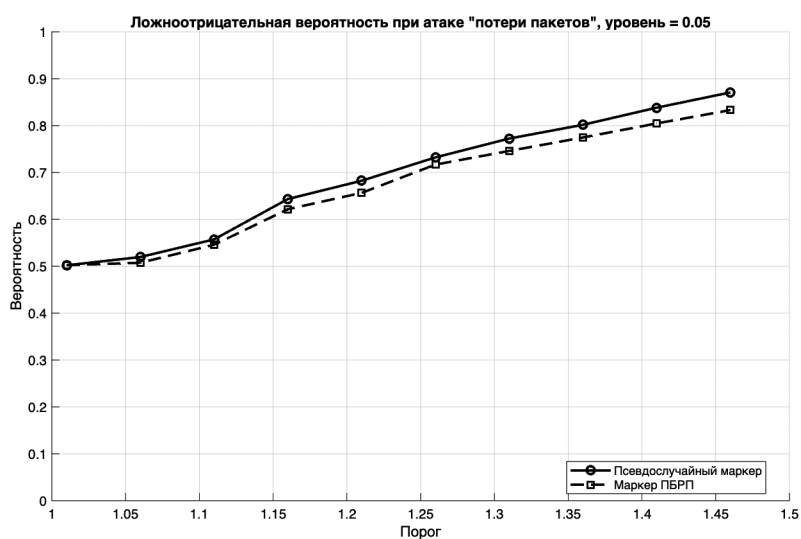
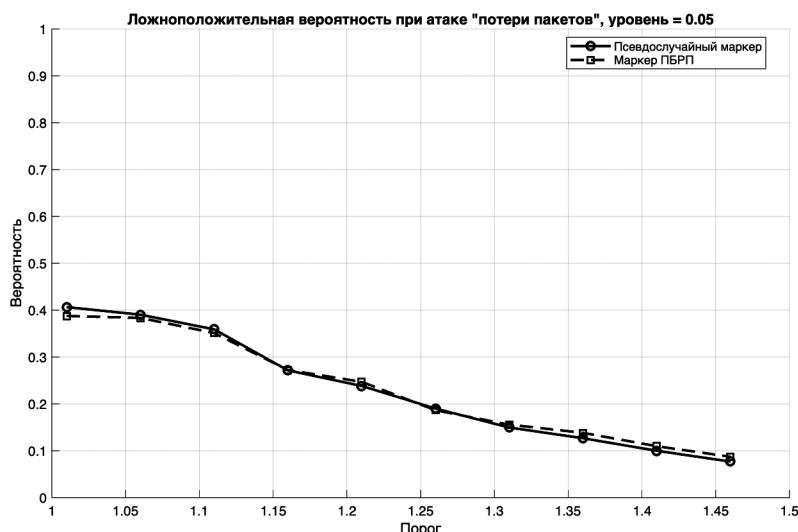


Рис. 4. Зависимость вероятности ложного срабатывания от порога τ при атаке временного джиттера

Рис. 5. Зависимость вероятности пропуска маркера от порога τ при атаке смешивания потоковРис. 6. Зависимость вероятности ложного срабатывания от порога τ при атаке смешивания потоковРис. 7. Зависимость вероятности пропуска маркера от порога τ при атаке потерь пакетов

Рис. 8. Зависимость вероятности ложного срабатывания от порога τ при атаке потерь пакетов

Анализ представленных зависимостей показывает, что наиболее приемлемая рабочая область соответствует значениям порога в окрестности $\tau \approx 1.31$. Ниже этого значения ложные срабатывания остаются относительно высокими, а при дальнейшем росте порога быстро возрастают пропуски, особенно при смешивании потоков и потерях пакетов. Для единообразного сравнения дальнейшие выводы относятся именно к этой рабочей области.

Анализ графиков показывает, что применение ПБРП в большинстве сценариев уменьшает вероятность пропуска маркера по сравнению с псевдослучайной последовательностью. Наиболее выраженное отличие наблюдается при джиттере: в рабочей области $\tau \approx 1.31$

вероятность пропуска у ПБРП заметно ниже при практически совпадающей вероятности ложного срабатывания.

При смешивании потоков ПБРП также даёт выигрыш по вероятности пропуска, хотя сопровождается небольшим увеличением вероятности ложного срабатывания. Для потерь пакетов преимущество сохраняется, но выражено слабее. При переупаковке различия по вероятности пропуска минимальны; основное достоинство ПБРП здесь связано со снижением ложных тревог.

Наиболее тяжёлым сценарием с точки зрения пропуска маркера являются потери пакетов, поскольку именно в этом случае вероятность ложноотрицательного решения достигает максимальных значений для обоих вари-

Таблица 1.

Сравнение характеристик детектирования в рабочей области $\tau \approx 1.31$

Вид воздействия	Вероятность пропуска, псевдослучайный маркер	Вероятность пропуска, ПБРП	Вероятность ложного срабатывания, псевдослучайный маркер	Вероятность ложного срабатывания, ПБРП	Интерпретация
Переупаковка пакетов	0.60	0.60	0.20	0.17	По вероятности пропуска различия минимальны; преимущество ПБРП проявляется прежде всего в снижении ложных срабатываний.
Джиттер	0.38	0.31	0.18	0.18	Наиболее заметный выигрыш ПБРП: вероятность пропуска ниже при практически том же уровне ложных срабатываний.
Смешивание потоков	0.57	0.50	0.16	0.18	ПБРП заметно уменьшает пропуски, но сопровождается небольшим ростом ложных тревог.
Потери пакетов	0.77	0.75	0.15	0.15	Различия между маркерами невелики; ПБРП немного уменьшает вероятность пропуска, не давая выраженного выигрыша по ложным срабатываниям.

Примечание — значения в таблице соответствуют точкам графиков в окрестности рабочей области $\tau \approx 1.31$.

антов маркера. По показателю ложноположительной вероятности наиболее неблагоприятной в выбранной рабочей области оказывается атака переупаковки пакетов. В совокупности это показывает, что применение ПБРП не устраняет влияние атак полностью, однако позволяет сделать процедуру обнаружения более устойчивой в типовых сценариях нарушения парной структуры интервалов, особенно при джиттере и смешивании потоков.

Заключение

В работе рассмотрен интервальный метод маркирования сетевого потока «лоскут», ориентированный на задачи выявления нарушителя в глобальной частично контролируемой сети. В отличие от подходов, основанных исключительно на временных задержках или корреляции отдельных интервалов, предлагаемый метод использует парную интервальную структуру, позволяющую формировать маркер в сетевом потоке

и восстанавливать его даже при наличии внешних искажений.

Результаты моделирования показывают, что при пороговом значении $\tau \approx 1.31$ метод сохраняет работоспособность при смешивании потоков, потерях пакетов, джиттере и переупаковке пакетов. Использование ПБРП в качестве опорной последовательности улучшает качество обнаружения по сравнению с псевдослучайным маркером, причём наиболее заметный эффект достигается при джиттере и смешивании потоков.

Полученные результаты показывают, что предложенный метод может использоваться в качестве основы для практических систем идентификации и трассировки нарушителя в условиях шумового и смешанного трафика. Перспективным направлением дальнейшей работы является уточнение рабочих порогов детектора и расширение набора атакующих сценариев.

ЛИТЕРАТУРА

1. Iacovazzi A., Elovici Y. Network Flow Watermarking: A Survey // IEEE Communications Surveys & Tutorials. 2017. Vol. 19, No. 1. P. 512–530. DOI: 10.1109/COMST.2016.2604405.
2. Zhang L., Kong Y., Guo Y., Wu Y., Wang Z. Survey on Network Flow Watermarking: Model, Interferences, Applications, Technologies and Security // IET Communications. 2018. Vol. 12, No. 14. P. 1639–1648. DOI: 10.1049/iet-com.2017.1053.
3. Iacovazzi A., Sarda S., Elovici Y. InFlow: Inverse Network Flow Watermarking for Detecting Hidden Servers // IEEE INFOCOM 2018 — IEEE Conference on Computer Communications. 2018. P. 747–755. DOI: 10.1109/INFOCOM.2018.8486375.
4. Yao Z., Zhang L., Ge J., Wu Y., Zhang X. An Invisible Flow Watermarking for Traffic Tracking: A Hidden Markov Model Approach // ICC 2019 — IEEE International Conference on Communications. 2019. P. 1–6. DOI: 10.1109/ICC.2019.8761135.
5. Hou J., Li Q., Tan R., Meng S., Zhang H., Zhang S. An Intrusion Tracking Watermarking Scheme // IEEE Access. 2019. Vol. 7. P. 141438–141455. DOI: 10.1109/ACCESS.2019.2943493.
6. Chen Y., Chen Y. A Network Flow Correlation Method Based on Chaos Theory and Principal Component Analysis // International Journal of Network Security. 2020. Vol. 22, No. 2. P. 242–249. DOI: 10.6633/IJNS.202003_22(2).07.
7. Yu L., Zhang L., Zhang Y., Wen W., Du X., Cao F. Dynamic Interval-based Watermarking for Tracking down Network Attacks // 2021 IEEE 21st International Conference on Software Quality, Reliability and Security (QRS). 2021. P. 52–61.
8. Yang K., Liu Z., Zeng Y., Ma J. Sliding Window Based ON/OFF Flow Watermarking on Tor // Computer Communications. 2022. Vol. 196. P. 66–75. DOI: 10.1016/j.comcom.2022.09.028.
9. Li T., Liu K., Feng W., Yang C., Luo X. HeteroTiC: A Robust Network Flow Watermarking Based on Heterogeneous Time Channels // Computer Networks. 2022. Vol. 219. Art. 109424. DOI: 10.1016/j.comnet.2022.109424.
10. Gao S., Zhang L. An Optimal LT Code for Network Flow Watermarking // Proc. SPIE 12506, Third International Conference on Computer Science and Communication Technology (ICCSCT 2022). 2022. Art. 1250614. DOI: 10.1117/12.2662345.
11. Гофман М.В., Корниенко А.А. Предельные биполярные последовательности для робастного маркирования цифровых аудиосигналов по методу лоскута // Информатика и автоматизация. 2023. Т. 22, № 2. С. 221–260. DOI: 10.15622/ia.22.2.1.
12. Cui J., Han K., Sha L. et al. An Efficient Hexadecimal Network Flow Watermark Method for Tracking Attack Traffic // Scientific Reports. 2023. Vol. 13. Art. 21111. DOI: 10.1038/s41598-023-48552-0.
13. Feng W., Luo X., Li T., Yang C. HSTW: A Robust Network Flow Watermarking Method Based on Hybrid Packet Sequence-Timing // Computers & Security. 2024. Vol. 139. Art. 103701. DOI: 10.1016/j.cose.2024.103701.
14. Feng W., Luo X., Li T., Yang C. IP-Peeling: A Robust Network Flow Watermarking Method Based on IP Packet Sequence // Chinese Journal of Electronics. 2024. Vol. 33, No. 3. P. 694–707. DOI: 10.23919/CJE.2022.00.366.
15. Молькова Л.Ю., Гофман М.В. Методы защиты маркированного сетевого потока с использованием технологии лоскута: анализ устойчивости и применения // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. 2025. № 6-2. С. 152–158. DOI: 10.37882/2223-2966.2025.06-2.25.
16. Molkova L.Y., Biliatdinov K.Z., Gluharev M.L. Modern Technologies for Marking Network Flows: From Classical Methods to Innovations // International Journal of Open Information Technologies. 2025. Vol. 13, No. 7. P. 22–26.
17. Yuan Y., Ge J., Cheng G. DeMarking: A Defense for Network Flow Watermarking in Real-Time // Computers & Security. 2025. Vol. 152. Art. 104355. DOI: 10.1016/j.cose.2025.104355.
18. Qiao S., Zhu H., Sha L., Wang M., Guo Q. DynMark: A Dynamic Packet Counting Watermarking Scheme for Robust Traffic Tracing in Network Flows // Computers & Security. 2025. Vol. 157. Art. 104571. DOI: 10.1016/j.cose.2025.104571.
19. Qiao S., Guo Q., Wang M., Zhu H., Rodrigues J.J.P.C., Lyu Z. Advances in Network Flow Watermarking: A Survey // Computers & Security. 2025. Vol. 159. Art. 104653. DOI: 10.1016/j.cose.2025.104653.

ГЛУБОКОЕ ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ: АЛГОРИТМЫ, ПРИЛОЖЕНИЯ И ВЫЗОВЫ

DEEP REINFORCEMENT LEARNING: ALGORITHMS, APPLICATIONS, AND CHALLENGES

Ngoula Brintch-Geoffrey

Summary. The article is devoted to the analysis of Deep Reinforcement Learning (DRL) methods. Particular attention is paid to the components of the basic architecture, key algorithms and techniques, as well as applications in fields such as robotics, games, and autonomous driving. Challenges and limitations of DRL are discussed, such as sample inefficiency and safety concerns. Promising directions for future research are identified, including meta-learning, hierarchical reinforcement learning, and the combination of DRL with formal methods. Recommendations for practitioners on the effective application of DRL are provided.

Keywords: deep learning, reinforcement learning, artificial intelligence, robotics, neural networks, algorithms, safety, scalability.

Нгула Бринч-Жофрей

аспирант, Тюменский индустриальный университет
brintchpetrovitch.galaktionov@mail.ru

Аннотация. Статья посвящена анализу методов глубокого обучения с подкреплением (DRL). Особое внимание уделяется компонентам базовой архитектуры, ключевым алгоритмам и техникам, а также приложениям в таких областях, как робототехника, игры и автономное вождение. Обсуждаются вызовы и ограничения DRL, такие как неэффективность выборки и проблемы безопасности. Идентифицируются перспективные направления для будущих исследований, включая мета-обучение, иерархическое обучение с подкреплением и комбинацию DRL с формальными методами. Предоставляются рекомендации для практиков по эффективному применению DRL.

Ключевые слова: глубокое обучение, обучение с подкреплением, искусственный интеллект, робототехника, нейронные сети, алгоритмы, безопасность, масштабируемость.

Введение

Область искусственного интеллекта (ИИ), известная как машинное обучение (ML), фокусируется на создании моделей и алгоритмов, способных обучаться на данных без явного программирования [5]. Техники ML позволяют компьютерам понимать паттерны и отношения в данных и использовать их для предсказаний или суждений о ранее неизвестных данных. Распознавание изображений и звуков, обработка естественного языка (NLP), системы рекомендаций, обнаружение мошенничества и предиктивное обслуживание являются примерами использования техник ML.

Сети искусственных нейронов с несколькими слоями взаимосвязанных узлов являются основой глубокого обучения (DL), подмножества ML, мотивированного структурой и функционированием человеческого мозга [3]. Модели DL могут автоматически обучаться представлениям сложных данных, постепенно извлекая высокоуровневые признаки из низкоуровневых. Алгоритмы DL продемонстрировали передовые производительности в различных дисциплинах, включая компьютерное зрение, NLP, распознавание речи и игры.

Дисциплина обучения с подкреплением (RL) фокусируется на обучении агентов тому, как принимать серию

решений в окружающей среде для максимизации кумулятивного сигнала вознаграждения. С момента своего первого введения в 1980-х годах, RL использовался для решения разнообразных проблем, включая игры, робототехнику и системы контроля [2]. Традиционные алгоритмы RL часто полагаются на ручные признаки и линейную аппроксимацию функций, что ограничивает их способность обрабатывать сложные задачи с пространствами состояний и действий высокой размерности.

Глубокое обучение с подкреплением (DRL) появилось как решение этой проблемы, используя глубокие нейронные сети (DNN) для представления политики и функции ценности агента. DNN могут автоматически обучаться иерархическим представлениям сложных входных данных, таких как изображения, звуки и естественный язык, что делает их хорошо подходящими для обработки пространств состояний и действий большой размерности [8]. Комбинация RL и DL привела к значительным продвижениям в области, где многие приложения достигли передовой производительности.

В последние годы наблюдается взрыв интереса к DRL, с многочисленными приложениями в промышленности и академической среде. Например, DRL использовался для обучения роботов выполнению задач, таких как захват объектов и навигация в сложных средах [7], игра в игры,

такие как Go и игры Atari, с сверхчеловеческой производительностью [15; 8], и генерация описаний естественного языка для изображений [9].

Литературный обзор

Проблема навигации и принятия решений автономных агентов изучается в рамках различных подходов. В контексте DRL, основные алгоритмы можно классифицировать на методы, основанные на ценности, и методы, основанные на политике.

Алгоритмы, основанные на ценности, такие как Deep Q-Networks (DQN), используют сеть для аппроксимации функции ценности действия. DQN был применен к широкому спектру задач, включая игры Atari [8] и контроль робототехнических систем [7]. Для повышения стабильности и сходимости метод DQN использует целевую сеть и память воспроизведения опыта.

Техники Актер-Критик (AC) используют две нейронные сети: сеть актера, которая обучает стохастическую политику, и сеть критика, которая измеряет функцию ценности [11]. Три подкатегории методов актер-критик существуют: актер-одиночка, актер-одиночка на политике и актер-одиночка вне политики. Техники на политике, такие как A2C и A3C, обновляют политику на основе текущей политики, в то время как техники вне политики, такие как DDPG и TD3, используют отдельную поведенческую политику [4].

Техники градиента политики (PG) оптимизируют непосредственно параметры политики для максимизации ожидаемого вознаграждения. Теорема PG предоставляет способ вычисления градиента ожидаемого вознаграждения относительно параметров политики [2]. Техники PG могут быть подразделены на техники VPG и техники TRPO. TRPO использует ограничение области доверия для гарантии того, что обновление политики не отклоняется слишком сильно от текущей политики.

PPO является алгоритмом RL на основе политики, разработанным в 2017 году. PPO использует усеченную суррогатную цель для обновления сети политики, что помогает избежать больших изменений политики, которые могли бы дестабилизировать процесс обучения [2].

SAC является другим алгоритмом RL на основе политики, который использует цель максимальной энтропии для поощрения исследования и стохастический актер для выборки действий [6]. TD3 является алгоритмом RL на основе политики, который использует две сети критика для оценки функции ценности состояния-действия и механизм отложенного обновления для стабилизации обучения [4].

Материалы и методы

В данном исследовании проводится систематический анализ алгоритмов DRL, представленный в литературе. Методология включает классификацию алгоритмов, анализ их архитектуры и оценку применимости к различным задачам.

Алгоритмы DRL комбинируют принципы RL с глубокими нейронными сетями для обучения сложных политик и функций ценности из входных данных высокой размерности. В работе рассматриваются наиболее широко используемые алгоритмы, такие как техники актер-критик, техники PG и сети глубокого Q-обучения (DQN).

Для оценки эффективности алгоритмов используются следующие критерии:

1. Эффективность выборки: Способность обучаться на меньшем количестве примеров.
2. Стабильность обучения: Сходимость алгоритма без значительных колебаний.
3. Производительность: Достижимое вознаграждение в тестовых средах.
4. Безопасность: Способность гарантировать безопасное поведение в критических приложениях.

Результаты

Приложения DRL

DRL продемонстрировал отличный потенциал в ряде приложений, от робототехники и игр.

Робототехника: Робототехника является одной из самых захватывающих областей, где может использоваться DRL. Агенты RL могут использоваться для контроля роботов, выполняющих сложные задачи в реальных средах. Например, RL использовалась для обучения роботов выполнению задач, таких как захват объектов, манипулирование инструментами и навигация в сложных средах [7]. Однако применение RL к робототехнике ставит несколько вызовов. Основным вызовом — потребность в эффективности выборки, так как реальные взаимодействия с роботом могут быть медленными и дорогостоящими. Для решения этого вызова исследователи разработали техники, такие как трансферное обучение и обучение на основе симуляции [9].

Игры: DRL также успешно применялся к играм. Агенты RL отличились в играх, таких как Go, Chess и других. Эти игры представляют уникальный вызов для агентов RL, так как они включают долгосрочное планирование и стратегическое мышление [15]. Одним из основных преимуществ использования RL для игр является то, что она может обучаться на самообучении, где агент играет против себя и улучшается со временем.

NLP: DRL также показал обещания в NLP. Агенты RL могут использоваться для обучения генерации текста, ответов на вопросы и выполнению других задач NLP. Например, RL использовалась для формирования чат-ботов, способных взаимодействовать с пользователями естественным [9].

Тестирование программного обеспечения: Исследования изучают, может ли RL заменить метаэвристические алгоритмы, разработанные человеком, в тестировании программного обеспечения на основе поиска (SBST). Новая модель GunPowder демонстрирует, что агент способен обучаться имитации метаэвристических алгоритмов SBST, достигая 100 % покрытия ветвей [10].

Сравнение алгоритмов

В таблице 1 представлены основные ограничения различных подходов к DRL.

Таблица 1.

Ограничения DRL

Ограничение	Описание
Неэффективность выборки	Алгоритмы DRL часто требуют больших количеств данных для обучения эффективных политик.
Проблемы безопасности	Агенты RL могут иногда обучаться эксплуатировать уязвимости среды неожиданным и потенциально опасным образом.
Масштабируемость	При увеличении числа состояний и действий алгоритмы могут становиться чрезмерно медленными.
Интерпретируемость	С усложнением агентов становится труднее понимать и интерпретировать их поведение.

Обсуждение

Ограничения и открытые вызовы

Несмотря на огромные продвижения, сделанные в DRL в последние годы, некоторые вопросы остаются, и некоторые ограничения еще не решены.

Неэффективность выборки является одним из основных недостатков DRL. Эффективные алгоритмы DRL часто требуют много данных для обучения, что может быть запретительным в практических приложениях. Необходимо разработать алгоритмы DRL, способные обучаться на меньшем количестве выборок, например, алгоритмы мета-обучения, которые могут обучаться на прошлом опыте [1].

Проблемы безопасности являются другим ограничением. Агенты RL могут иногда обучаться эксплуатировать среду неожиданным и потенциально опасным образом, что вызывает опасения по поводу безопасности.

Необходимо разработать алгоритмы DRL, способные гарантировать безопасное и надежное поведение, например, алгоритмы, которые интегрируют ограничения безопасности в процесс обучения [12].

Масштабируемость также является major проблемой в DRL. Традиционные алгоритмы DRL могут становиться чрезмерно медленными или требовать значительных ресурсов памяти по мере увеличения числа состояний и действий в среде. Необходимо разработать алгоритмы DRL, способные адаптироваться к более обширным и сложным средам, например, алгоритмы иерархического RL, способные обучаться на нескольких уровнях абстракции [1].

Интерпретируемость и объяснимость являются важными вопросами в DRL. По мере того как агенты RL становятся более сложными и мощными, становится все более трудным понимать и интерпретировать их поведение. Необходимо разработать алгоритмы DRL, способные предоставлять объяснения для своих решений и действий [13].

Комбинация с формальными методами

Формальные методы являются математическими техниками для спецификации, анализа и верификации программных и аппаратных систем [13; 14]. Они часто используются для гарантии корректности и безопасности критических систем, таких как используемые в аэрокосмической отрасли, медицинских устройствах и автономных транспортных средствах [12].

Алгоритмы DRL могут иногда обучаться политикам, которые нарушают ограничения безопасности. Это особенно беспокоит в критически важных для безопасности приложениях (safety-critical приложениях). Для решения этой проблемы растет интерес к комбинации DRL и формальных методов [12; 13]. Идея состоит в использовании формальных методов для спецификации ограничений безопасности и требований, а затем использовании DRL для обучения политик, которые удовлетворяют этим ограничениям и требованиям [14].

Заключение

В статье представлен обзор DRL, его приложений, ограничений и открытых проблем, с акцентом на некоторые из самых перспективных областей исследований в этой области. Обсуждены алгоритмы DRL, такие как сети глубокого Q-обучения, техники актер-критик и техники градиента политики [2; 8; 11].

Обсуждены продвижения в мета-обучении, иерархическом RL и обучении по имитации. Обсуждены приложения DRL в различных областях, включая робото-

технику [7], игры [15], NLP [9] и системы рекомендаций. Исследована комбинация DRL с другими техниками, такими как трансферное обучение и многозадачное обучение, для улучшения эффективности и действенности DRL [1].

Для помощи практикам в эффективном применении DRL в их работе предоставлены рекомендации, которые включают советы по началу с простых проблем, выбору соответствующих алгоритмов и архитектур, обращению внимания на безопасность и этику, сотрудничеству с экспертами и с последними исследованиями в области [1].

В целом, DRL является областью, находящейся в состоянии постоянного развития, с огромным потенциалом влияния на широкий спектр приложений. По мере того как эта сфера продолжает эволюционировать, вероятно появление новых методов и подходов для продвижения передовых достижений и освоения новых областей применения DRL. Однако важно отметить: по мере того как алгоритмы DRL становятся более мощными и сложными, необходимо гарантировать их безопасность, надежность и прозрачность [12].

ЛИТЕРАТУРА

1. Krichen M. Apprentissage par Renforcement Profond. HAL Id: hal-04139754. 2022.
2. Sutton R.S., Barto A.G. Reinforcement Learning: An Introduction. MIT Press, 2018.
3. Boulila W., Driss M., Alshanqiti E., et al. Weight initialization techniques for deep learning algorithms in remote sensing. Advances on Smart and Soft Computing. 2022. P. 477–484.
4. Fujimoto S., Hoof H., Meger D. Addressing function approximation error in actor-critic methods. International conference on machine learning. 2018. P. 1587–1596.
5. Cedric L.S., Adoni W.Y.H., Aworka R., et al. Crops yield prediction based on machine learning models. Smart Agricultural Technology. 2022. P. 100049.
6. Haarnoja T., Zhou A., Abbeel P., Levine S. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning. International conference on machine learning. 2018. P. 1861–1870.
7. Gu S., Holly E., Lillicrap T., Levine S. Deep reinforcement learning for robotic manipulation. IEEE ICRA. 2017. P. 3389–3396.
8. Mnih V., Kavukcuoglu K., Silver D., et al. Human-level control through deep reinforcement learning. Nature. 2015. Vol. 518. P. 529–533.
9. Hafner D., Lillicrap T., Fischer I., et al. Learning latent dynamics for planning from pixels. Int. conference on machine learning. 2019. P. 2555–2565.
10. Kim J., Kwon M., Yoo S. Generating test input with deep reinforcement learning. Proceedings of the 11th International Workshop on Search-Based Software Testing. 2018. P. 51–58.
11. Konda V., Tsitsiklis J. Actor-critic algorithms. Advances in neural information processing systems. 1999. Vol. 12.
12. Krichen M. Model-based testing for real-time systems. PhD thesis. 2007.
13. Krichen M. A formal framework for conformance testing of distributed real-time systems. International Conference On Principles Of Distributed Systems. 2010. P. 139–142.
14. Krichen M. Contributions to model-based testing of dynamic and distributed real-time systems. PhD thesis. 2018.
15. Silver D., Huang A., Maddison C.J., et al. Mastering the game of go with deep neural networks and tree search. Nature. 2016. Vol. 529. P. 484–489.

© Нгула Бринч-Жофрей (brintchpetrovitch.galaktionov@mail.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

НЕЙРОННАЯ СЕТЬ ДЛЯ ПЛАНИРОВАНИЯ ПЕРЕМЕЩЕНИЯ АВТОНОМНОГО РОБОТА

NEURAL NETWORK FOR PATH PLANNING OF AN AUTONOMOUS ROBOT

**D. Petrov
M. Petrov**

Summary. This paper discusses the development and implementation of a neural network for planning the movement of mobile robots in an autonomous navigation task. Existing approaches to this task are limited in the presence of moving obstacles in the target zone. The primary goal of this study is to develop an effective tool for planning the movement of a mobile robot in an environment with moving obstacles. A deep reinforcement learning approach can improve performance in such conditions. This paper briefly reviews the main areas of deep reinforcement learning, presents the neural network architecture, training using the DDPG algorithm, and experimental verification of its effectiveness. A key feature of the presented architecture is a specialized structure for processing lidar data over multiple frames, capable of accounting for the dynamic properties of the environment using convolution layers and LSTM, extracting spatial and temporal features. The obtained results demonstrate sufficient effectiveness in a dynamic environment and potential for improvement.

Keywords: path planning, mobile robotics, neural networks, machine learning, reinforcement learning, deep learning.

Петров Денис Сергеевич

аспирант, Чувашский государственный университет
имени И.Н. Ульянова, г. Чебоксары
amitie2104@gmail.com

Петров Максим Сергеевич

Чувашский государственный университет
имени И.Н. Ульянова, г. Чебоксары
petrovms_2023@mail.ru

Аннотация. В данной статье рассматривается разработка и реализация нейронной сети для планирования перемещения средств мобильной робототехники в задаче автономной навигации. Существующие подходы к решению данной задачи ограничены при наличии подвижных препятствий в целевой зоне. Основной целью исследования является разработка эффективного средства планирования перемещения мобильного робота в среде с подвижными препятствиями. Подход на основе глубокого обучения с подкреплением может повысить эффективность работы в подобных условиях. Кратко рассмотрены основные направления глубокого обучения с подкреплением, представлена архитектура нейронной сети, обучение с использованием алгоритма DDPG и экспериментальная проверка эффективности. Ключевой особенностью представленной архитектуры является специальная структура для обработки данных лидара за несколько кадров, способная учитывать динамические свойства окружения, используя слои свертки и LSTM, путем извлечения пространственных и временных признаков. Полученные результаты показывают достаточную эффективность в условиях динамичного окружения и потенциал для улучшения.

Ключевые слова: планирование перемещения, мобильная робототехника, нейронные сети, машинное обучение, обучение с подкреплением, глубокое обучение.

Введение

В современных условиях всё более актуальной становится задача создания робототехнических систем, способных к полностью автономному передвижению. Ключевыми критериями таких систем являются безопасность и эффективность траекторий движения. Для их обеспечения необходимо принимать во внимание как стационарные, так и подвижные препятствия, находящиеся в рабочей зоне робота. Наиболее значимым остаётся вопрос планирования маршрута в условиях неопределённой среды, где информация о препятствиях может быть неполной или изменчивой.

Поскольку рассматривается локальная часть задачи планирования перемещения, требуется разработать локальный планировщик, который бы гарантировал безопасный и оптимальный переход мобильного робота к за-

данной целевой точке, учитывая текущие ограничения пространства и динамику окружающих объектов.

В общем случае, в качестве локального метода планирования может применяться нейронная сеть, обученная для решения задачи избегания препятствий и точной навигации в пределах выбранного маршрута. Нейронная сеть с подходящей архитектурой может использовать информацию о текущем состоянии мобильного робота, включая данные с сенсоров (например, LIDAR, ультразвуковой сенсор, камеры). Такой подход позволяет управляемому объекту принимать адаптивные решения в режиме реального времени, эффективно взаимодействуя с неожиданными изменениями в окружающей среде.

Для этого стоит использовать нейронные сети с подкреплением, как средство способное на самообучение в различных условиях. Исходя из результатов, полу-

ченных в [1], наиболее перспективным подходом к эффективному планированию пути является метод глубокого обучения с подкреплением, способный учитывать данные об окружении и эффективно использовать данные о состоянии робота, такие как показания сенсоров и конфигурация объекта в пространстве.

Существует несколько основных направлений применения глубокого обучения с подкреплением в задачах планирования перемещения.

Теория

Для описания задачи обучения с подкреплением необходимо определить ключевые компоненты, которые часто формализуются как Марковский процесс принятия решений (МППР), общая схема взаимосвязи на рисунке 1.

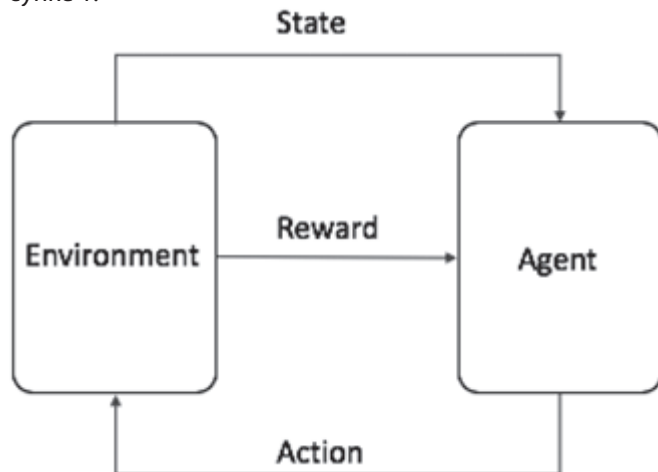


Рис. 1. Марковский процесс принятия решений

Требуется описать Агента (Agent), который получает наблюдаемое состояние (State), принимает действия (Action), воздействующие на среду (Environment). В зависимости от награды (Reward) обучается определенной стратегии поведения. Все эти компоненты будут последовательно определены в дальнейшей работе.

Глубокое обучение с подкреплением можно разделить на два основных подхода: поиск оптимальной функции ценности и прямую оптимизацию стратегии движения.

В методах, основанных на функции ценности, например DQN[2], сеть оценивает ожидаемую выгоду каждой пары «состояние-действие», после чего выбирается действие с наибольшим Q-значением. Такие алгоритмы хорошо работают при дискретных управляющих сигналах, но применение к реальному объекту непрактично и сопряжено с потерей точности.

Политико-ориентированные схемы, к примеру PPO[3], параметризуют непосредственно политику

и корректируют её градиентом ожидаемого вознаграждения. Их преимущество — естественная работа с непрерывными действиями, однако без вспомогательной оценки ценности обучение может быть шумным и медленным.

Современные методы глубокого обучения часто объединяют оба подхода. Сеть-«Актор» (Actor) отвечает за стратегию (политику), а сеть-«Критик» (Critic) оценивает действия актера (функция ценности), что позволяет ускорить обучение и повысить его стабильность. Именно поэтому гибридные архитектуры, сочетающие актора и критика, стали стандартом: актор генерирует действие, критик оценивает его качество, а ошибка оценки служит сигналом для обновления обеих сетей.

На сегодняшний день самым перспективным представителем гибридных методов считается Deep Deterministic Policy Gradient (DDPG), подробно описанный в [4]. Благодаря буферу опыта и целевым сетям DDPG стабилизирует обучение даже в условиях высокой шумности сенсоров, что делает его предпочтительным выбором для задач мобильной робототехники с непрерывным управлением, зарекомендовавшим себя при решении таких задач как следование за подвижным объектом [5], отслеживание траектории колесного робота [6], движения БПЛА в городской местности [7].

Это делает DDPG предпочтительным выбором для обучения собственного алгоритма планирования перемещения.

Исходя из изложенного в введении, в минимальный набор информации, передаваемой нейронной сети для генерации команды движения должны входить следующие данные:

x, y, z — координаты робота в пространстве

v_x, v_y, v_z — значение линейной скорости робота

w_x, w_y, w_z — значение угловой скорости робота

φ_a — курсовой угол

D_g — расстояние до цели

st_t — шаг в эпизоде, в будущем может быть коэффициентом временных или энергетических затрат

$\{l_1 \dots l_n\}$ — показания лидара за последние несколько кадров, для оценки динамики окружения.

В результате нейронная сеть должна сгенерировать желаемое значение линейного v_r и углового ускорения w_r , поскольку это даст более плавное движение и учет инерции, чем мгновенное изменение скорости движе-

ния или поворота. Для моделирования движения агента используется динамическая модель робота с дифференциальным приводом. Управляющими воздействиями являются линейное и угловое ускорения, что позволяет учитывать инерционные характеристики робота при маневрировании среди подвижных препятствий. Максимальная линейная скорость робота — 0.25 м/с, максимальная угловая скорость — 2.5 рад/с, для плавности движения и стабильности показаний — максимальное линейное ускорение робота — 0.2 м/с², максимальная угловая скорость — 1.0 рад/с².

Таким образом, агент — DDPG, вектор состояния $s_t \in S = \{x, y, z, v_x, v_y, v_z, w_x, w_y, w_z, \phi, D_g, st, \{l_1 \dots l_n\}\}$, вектор действий $a_t \in A = \{v, w\}$. Среда, награда за взаимодействие и то, как агент обучается в ней будут описаны далее.

Архитектура нейронной сети

Особенность архитектуры актора состоит в том, что данные о состоянии робота и данные лидача на первом шаге обрабатываются отдельно. Данные о состоянии робота проходят через 3 последовательных полносвязных слоя с функцией активации ReLU. Данные лидача, которые состоят из сканов за последние 3 кадра, проходят через 2 последовательных слоя свертки для фильтрации параметров и слой LSTM, который позволит отследить динамику окружения, подобная структура использовалась в [8] для определения характера действий людей. Затем обработанные данные соединяются для обработки в полносвязном слое и поступают в отдельные полносвязные слои для генерации 2 разных команд движения. Архитектура актора на рисунке 2, а.

Архитектура критика частично повторяет архитектуру актора для понимания динамических свойств, после их выделения данные проходят через 4 последовательных полносвязных слоя для генерации общей оценки, архитектура критика на рисунке 2, б.

Начисление наград происходит следующим образом, таблица 1.

Таблица 1.

Награды и условия

Условие	Значение
Робот достиг цели	200
Столкновение или выход за предел	-100
Приближение к цели	$10 * (dict_{t-1} - d_t)$
Штраф за каждый шаг	$-0.1 * (1 + step/1000)$

Виды и значения наград определялись путем создания необходимого минимума: алгоритм должен стре-

миться к цели и избегать столкновений, увеличенная награда за достижение перекроет штраф за возможное уклонение в сторону и задержку, для борьбы с круговым движением и «накруткой» награды в ходе обучения был добавлен штраф за выживание [9], скорректированный с учетом номера шага для прогрессирующего эффекта, как следствие, это должно уменьшить затраты на движение. При столкновении и выходе за предел карты не только назначается штраф, но и завершается эпизод обучения, это должно способствовать соблюдению безопасности в будущем.

В качестве среды обучения выступает симулированная двумерная среда, представляющая собой пространство 10x10 метров. В начале каждого эпизода позиция робота и цели задавалась случайным образом, с минимальным расстоянием между ними в 4 метра. Генерация препятствий, количество и тип зависят от сценария обучения. При генерации проводится проверка на достижимость цели и минимальный свободный радиус вокруг робота, для возможности уклонения. Размер буфера опыта равен 10^6 , как промежуточный вариант между слишком коротким лимитом опыта и раздутием затрат памяти. Максимальное число шагов в эпизоде — 5000, для лучшего обучения в начале и возможности накопления опыта. Данные шума и гиперпараметры взяты из [4], в данной статье проверяется эффективность архитектуры в среде с подвижными препятствиями.

Обучение нейросети происходило с последовательно возрастающей сложностью (Таблица 2).

Таблица 2.

Процесс обучение нейросети

Уровень сложности	Количество эпизодов	Состав препятствий	Среднее число шагов
Начальный	800	3 стационарных	726.3
Средний	1200	4 стационарных и 1 подвижное	601.2
Повышенный	2000	5 стационарных и 3 подвижных	691.4

На начальном уровне сложности алгоритм усваивает базовые правила перемещения, на среднем учитывает возможность наличия подвижных препятствий и на повышенном обучается уходу от столкновений. Число шагов ожидаемо выше при начальном обучении с нуля, уменьшается в процессе и возрастает при повышении сложности. Подвижные препятствия имеют скорость в половину максимальной скорости робота.

Для проверки эффективности обученной модели было проведено тестирование на наборе из 200 эпизодов, робот достиг цели в 82 % случаев, что является хорошим показателем при учете ограниченного набора

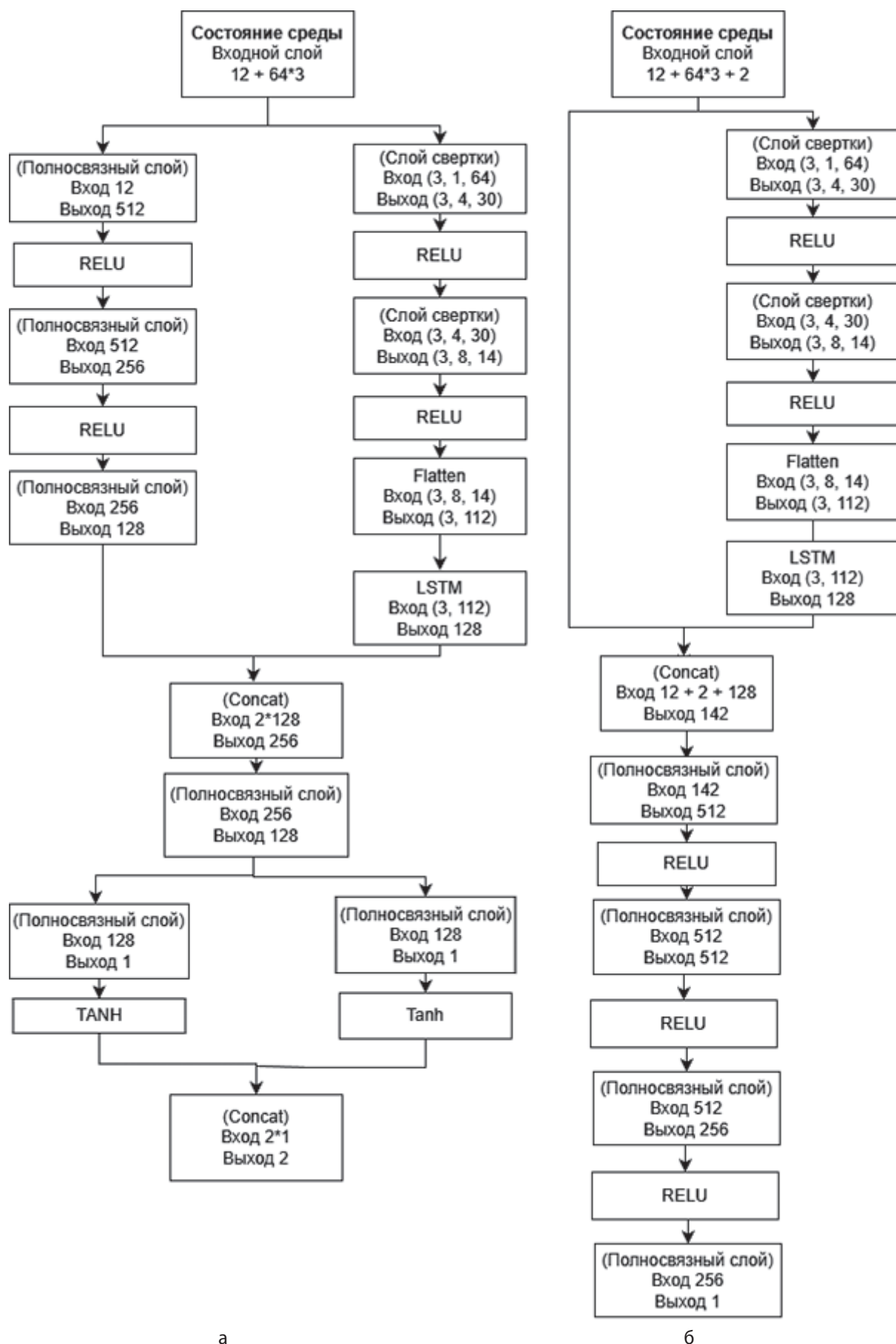


Рис. 2. Архитектура сети: а — актер, б — критик

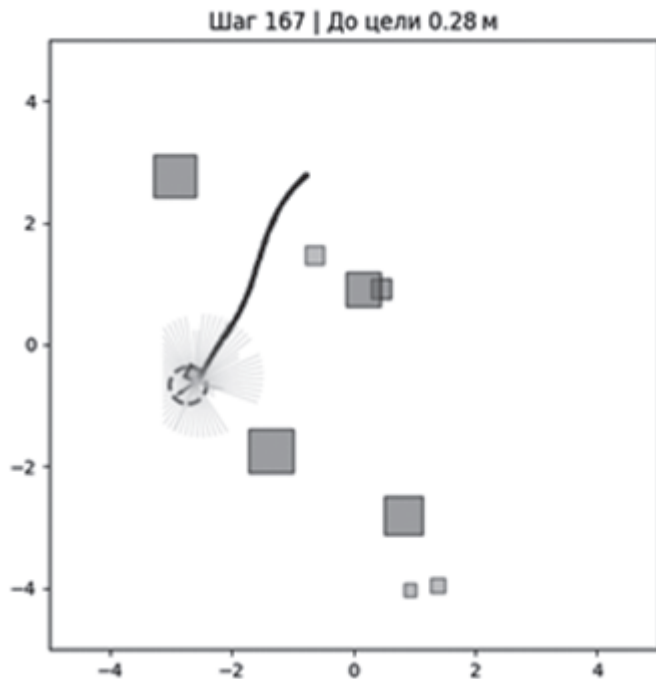


Рис. 3. Тестовый эпизод

обучающих данных и случайном характере движения препятствий. Пример траектории движения робота в одном из тестовых эпизодов на рисунке 3.

Как видно, робот предпринимал маневры уклонения от одного из подвижных препятствий, затем продолжал движение максимально прямо к цели, что соответствует требованиям к безопасному и эффективному перемещению в среде с подвижными препятствиями.

Заключение

Таким образом, глубокое обучение с подкреплением посредством DDPG может быть использовано для реализации локального планирования перемещения с учетом данных сенсоров и наличия подвижных препятствий в зоне движения. Экспериментально доказана эффективность применения подобного подхода. В ходе дальнейших работ ожидается уточнение гиперпараметров и функции награды для лучшего уклонения от препятствий, включая учет курсового угла и расстояние до ближайших объектов, дообучение в среде Gazebo и создание контроллера физического мобильного робота в ROS2.

ЛИТЕРАТУРА

1. Петров Д.С. Исследование применения нейронных сетей для планирования перемещения автономного робота // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и Технические Науки. — 2026. — №02. — С. 128–134 DOI 10.37882/2223-2966.2026.02.28
2. Mnih V. et al. Playing atari with deep reinforcement learning — arXiv: 1312.5602. [Preprint]. 2013. — URL: <https://arxiv.org/abs/1312.5602>
3. Schulman J., Wolski F., Dhariwal P. et al. Proximal policy optimization algorithms. — arXiv:1707.06347 [Preprint]. — 2017. — URL: <https://arxiv.org/abs/1707.06347>
4. Lillicrap T.P., Hunt J.J., Pritzel A., Heess N., Erez T., Tassa Y., Silver D., Wierstra D. Continuous control with deep reinforcement learning. — arXiv: 1509.02971 [Preprint]. — 2015. — URL: <http://arxiv.org/abs/1509.02971>
5. Звонарев И.С., Караваев Ю.Л. Нейросетевой алгоритм обучения мобильного робота в задаче следования за целью //Вестник ИжГТУ имени М.Т. Калашникова. — 2024. — Т. 27. — №. 2. — С. 4–14.
6. Gou W., Liu Y. Trajectory tracking control of wheeled mobile robot based on improved LSTM-DDPG algorithm //Journal of Physics: Conference Series. — IOP Publishing, 2022. — Т. 2303. — №. 1. — С. 012069.
7. Tian S. et al. Fast UAV path planning in urban environments based on three-step experience buffer sampling DDPG //Digital Communications and Networks. — 2024. — Т. 10. — №. 4. — С. 813–826.
8. Zhou H. et al. Efficient human activity recognition on edge devices using DeepConv LSTM architectures //Scientific Reports. — 2025. — Т. 15. — №. 1. — С. 13830.
9. Jeng S.L., Chiang C. End-to-End Autonomous Navigation Based on Deep Reinforcement Learning with a Survival Penalty Function //Sensors. — 2023. — Т. 23. — №. 20. — С. 8651.

© Петров Денис Сергеевич (amitie2104@gmail.com); Петров Максим Сергеевич (petrovms_2023@mail.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

ПРИНЦИП ЧЕЛОВЕКО-МАШИННОГО КОНСЕНСУСА ПРИ ИНТЕЛЛЕКТУАЛЬНОЙ ПОДДЕРЖКЕ ПРИНЯТИЯ РЕШЕНИЙ В АВТОМАТИЗИРОВАННЫХ СИСТЕМАХ НЕЛИНЕЙНОГО УПРАВЛЕНИЯ ДОСТАВКОЙ ИНСУЛИНА

THE PRINCIPLE OF HUMAN-MACHINE CONSENSUS IN INTELLIGENT DECISION SUPPORT IN AUTOMATED INSULIN DELIVERY NON-LINEAR CONTROL SYSTEMS

K. Pozhar

Summary. Developing technologies for creating next-generation personalized, automated medical systems for insulin delivery control based on complex predictive models requires a new approach to organizing human-machine interactions. The paper proposes the principle of human-machine consensus, which states that significant control decisions are executed only with the consensus of the system and the user. A significant result of the transition from a hierarchical to a synergistic interaction structure is the ability to implement non-interpretable by human nonlinear computational models, which significantly improves the effectiveness of therapy. Algorithms for implementing control and decision support systems based on this principle are proposed.

Keywords: insulin therapy, automation, nonlinear control, decision support, human-machine consensus, diabetes mellitus.

Пожар Кирилл Витольдович

Кандидат технических наук, доцент, Национальный
исследовательский университет МИЭТ (Москва)
pozhar@bms.zone

Аннотация. Разработка технологий создания персонализированных автоматизированных медицинских систем управления доставкой инсулина нового поколения на основе сложных прогнозирующих моделей требует нового подхода к организации взаимодействия человек-машина. В работе предлагается принцип человеко-машинного консенсуса, заключающийся в том, что значимые решения по управлению исполняются только при консенсусе системы и пользователя. Качественным результатом перехода от иерархической структуры взаимодействия к синергетической является возможность внедрения неинтерпретируемых пользователем нелинейных расчётных моделей, что позволяет значительно повысить эффективность терапии. Предложены алгоритмы реализации систем управления и поддержки принятия решений на основе предложенного принципа.

Ключевые слова: инсулинотерапия, автоматизация, нелинейное управление, поддержка принятия решений, человеко-машинный консенсус, сахарный диабет.

Введение

Среди актуальных задач создания медицинских изделий и систем нового поколения одной из наиболее важных является создание автоматизированной системы управления доставкой инсулина для пациентов с сахарным диабетом 1-го типа. Она характеризуется высокой социальной значимостью ввиду высокой распространённости сахарного диабета 1-го типа среди молодых людей трудоспособного возраста. При этом неэффективная терапия данного заболевания приводит к ранней инвалидизации и трудовым потерям для экономики.

В свою очередь низкая эффективность терапии зачастую обусловлена чрезмерной нагрузкой на пользователя, который вынужден непрерывно совмещать роли объекта управления, оператора системы управления, лица, принимающего решения и системы биологической обратной связи.

Основные научно-технические тренды в автоматизации систем управления доставкой инсулина при диабете 1-го типа связаны с переходом от ручного управления инсулиновой помпой по данным непрерывного мониторинга глюкозы, при котором вычисление параметров инфузии инсулина осуществляется пользователем по арифметическим формулам или с использованием простых болюс-калькуляторов, к системам с автоматическим расчётом управляющих воздействий [1], использующих нелинейные модели естественной системы регуляции глюкозы в организме [2] в сочетании с мощным аппаратом оптимизации [3]. Однако сложность таких методов и алгоритмов уже не позволяет пользователю понимать и контролировать работу системы, поскольку человеку несвойственны нелинейные оценки, что может дополнительно затруднить взаимодействие человека с технической системой.

Другой тренд связан с развитием телемедицины, персональных медицинских помощников, направлен-

ных на повышение связности процесса терапии за счёт включения медицинских специалистов, а также поддержку принятия решений пользователем. В условиях развития возможностей современных вычислительных и коммуникационных систем указанные тренды ставят вопрос о выходе систем инсулинотерапии на новое качество взаимодействия человек-машина, основанное на сочетании возможностей технических систем к прогнозированию нелинейных процессов и оптимальному управлению сложными системами и доступных пользователю возможностей анализа контекста, обеспечивающее, несмотря на общий рост сложности, упрощение взаимодействия.

В настоящей работе обсуждается новый подход к поддержке принятия решений, необходимый для внедрения автоматизированных систем нелинейного управления доставкой инсулина.

Поддержка принятия решений в системах управления доставкой инсулина

Объектом управления является система регуляции глюкозы в организме человека, которая представляет собой сложную биологическую систему с большим количеством внутренних и межсистемных связей, многие из которых являются не полностью изученными или описанными.

Рассматриваемые системы управления доставкой инсулина основаны на комбинации линейного или нелинейного управления по обратной связи и прогнозирующего управления [4, 5]. Управление по обратной связи характеризуется высокой инерционностью, что не позволяет эффективно компенсировать возмущения высокой амплитуды. Прогнозирующее управление основано на оптимизации управляющих воздействий с использованием прогнозирующих математических моделей объекта управления, заданных в виде нелинейных систем ОДУ. Как правило прогнозирующее управление применяется при приёме пациентом пищи и других внешних возмущениях достаточно высокой амплитуды. Управление по обратной связи на это время отключается или ограничивается по амплитуде и функционирует полноценно вдали от высокоамплитудных внешних возмущений.

Обеспечение системы управления информацией осуществляется пациентом через интерфейс, например, смартфон, посредством которого он сообщает состав принятой пищи и планируемое время приёма или параметры других значимых событий [6].

Помимо этого, существующие методы управления подразумевают задействование пользователем собственных сенсорных и когнитивных возможностей для оценки плохо-формализуемых воздействий и состояний

[7]. Ограниченность информации в системе управления и высокая зависимость от точности вводимой пользователем информации вынуждает ограничивать функционал технических систем формированием рекомендаций, которые могут утверждаться или отклоняться пользователем. В качестве альтернативы пользователь может ввести своё решение, что требует от него учитывать и обрабатывать широкий спектр информации.

Данная подчинённая структура взаимодействия человек-машина оставляет полную ответственность за принятие решений по управлению на пользователе, что либо не позволяет использовать алгоритмы вычисления управляющих воздействий отличные от доступных пониманию пользователя, либо требует от пользователя поиска компромисса между доверием к вычислениям технической системы и риском допустить ошибку в собственных оценках.

Принцип человеко-машинного консенсуса

В работе мы предлагаем новый базовый принцип управления: значимые решения по управлению исполняются только при консенсусе системы и пользователя.

В его основе остаётся полагание, что техническая система может допускать ошибки в расчётах ввиду недостаточной информированности и любое решение должно быть верифицировано пользователем. В то же время к нему добавляется утверждение, что высокопроизводительная техническая система, основанная на точных прогнозирующих моделях, при наличии достаточной информации позволяет вырабатывать значительно более эффективные управляющие воздействия, оптимизировать их сложные формы, работать с нелинейными моделями, сложными для умозрительной оценки пользователем. Из этого следует ряд выводов:

- решения технической системы являются предпочтительными;
- пользователь имеет право их отклонять и предлагать свои решения;
- предлагаемые пользователем решения должны проходить оценку эффективности и безопасности.

Последнее особенно актуально при плохом самочувствии пациента, когда вероятность ошибочной оценки значительно возрастает.

Данный принцип обеспечивает взаимную проверку, обоснование и согласование решений, обеспечивая переход от контроля решений через контекст к синергии возможностей вычислительной системы и человека.

Алгоритмы поддержки принятия решений

На основании предложенного принципа предложен базовый подход к построению алгоритма взаимодействия человек-машина, включающий следующие блоки:

- выработка и согласование с пользователем технической системой оптимальных решений;
- выработка и согласование с пользователем технической системой альтернатив;
- анализ решений пользователя;
- проверка способности пользователя принимать решения;
- выработка рекомендаций для пользователя.

При отсутствии значительных отклонений в концентрации глюкозы от целевого диапазона, управление осуществляется по обратной связи без необходимости принятия решений. Выработка технической системой решений происходит либо по поступлении информации от пользователя, либо при выявлении расхождений прогнозируемых и измеренных значений или их трендов. Причиной расхождений могут быть ошибки пользователя в информировании системы, неточность прогноза в связи с флуктуациями в объекте управления или неисправность датчика обратной связи или инсулинового насоса. Процесс диалога системы поддержки принятия решений с пользователем направлен на выявления причин расхождений и выработка рекомендаций пользователю для их устранения: ввод недостающих данных, выполнение ранее заявленного, но пропущенного приёма пищи, калибровка монитора глюкозы, проверка целостности магистралей насоса и др. При этом, поскольку допускается недостаток информации в системе, и система переключается в режим более безопасного, но менее эффективного управления. В случае незначительных расхождений может осуществляться подстройка параметров звена управления по обратной связи.

Далее осуществляется проверка необходимости управляющих воздействий. Она основана на сравнении прогнозируемой эффективности управления без принятия дополнительных управленческих решений и с принятием таковых. Если эффективность управления в итоге растёт, или остаётся высокой, или будучи достаточно высокой падает незначительно, то необходимость в расчёте дополнительного управляющего воздействия отсутствует.

Если принято решение о необходимости управляющего воздействия, происходит оптимизация амплитудно-временного распределения инсулина. Алгоритм глобальной оптимизации предлагает набор параметров воздействия, который локально оптимизируется, а среди найденных локальных оптимумов выбирается наиболее эффективный. Если полученная эффективность слишком низкая, ситуация признаётся опасной и предлагается воспользоваться рекомендациями врача, хранящимися в памяти системы.

Предложенные оптимизированные параметры предлагаются пользователю для подтверждения. Важ-

но отметить, что при исполнении такого управляющего воздействия звено непрерывного автоматического управления переходит в режим программного управления с минимальной амплитудой. Предлагаемое решение направляется на согласование пользователю с визуализацией предполагаемой динамики глюкозы, что облегчает пользователю понимание потенциального эффекта. Параллельно с этим осуществляется расчёт альтернативных решений другой амплитуды и формы.

В случае несогласия пользователя на оптимальное и альтернативные решение и ни на одну из альтернатив, ему предлагается самостоятельно задать параметры управляющего воздействия на основе информации, которой он владеет и медицинских рекомендаций.

Предложенное пользователем решение исполняется только при подтверждении его безопасности по прогнозу. В противном случае вопрос способности пользователя принимать решения нуждаются в подтверждении, поскольку его решения могут быть оправданы, например плохим самочувствием или вмешательством извне. Аналогичная проверка происходит в случае отказа пользователя от всех предложенных альтернатив и от самостоятельного принятия решений. Она включает проверку когнитивного состояния пользователя посредством, например, теста Тьюринга, и удостоверение личности пользователя, например, посредством секретного вопроса. Данный подход обеспечивает взаимную верификацию решений, снимая с пользователя часть ответственности за принятие решений.

По результатам диалога либо осуществляется исполнение согласованного системой и пользователем решения, либо вырабатываются рекомендации для пользователя, а система переходит в режим слабого управления до устранения причин.

На рис. 1 представлен общий верхнеуровневый алгоритм работы системы управления и принятия решений, включающий основные реакции системы на возникающие ситуации.

Обсуждение

Роль системы поддержки принятия решений по управлению концентрацией глюкозы в крови при сложном многоуровневом адаптивном нелинейном управлении заключается в обеспечении консенсуса человека и машины в ситуациях, когда необходим синтез формализованных данных с высокой вычислительной сложностью технической системы и неформализованных данных, доступных пользователю. Для эффективного участия в это процессе пользователь должен быть в состоянии уделить внимание человеко-машинному диалогу. В моменты, когда это по тем или иным причи-

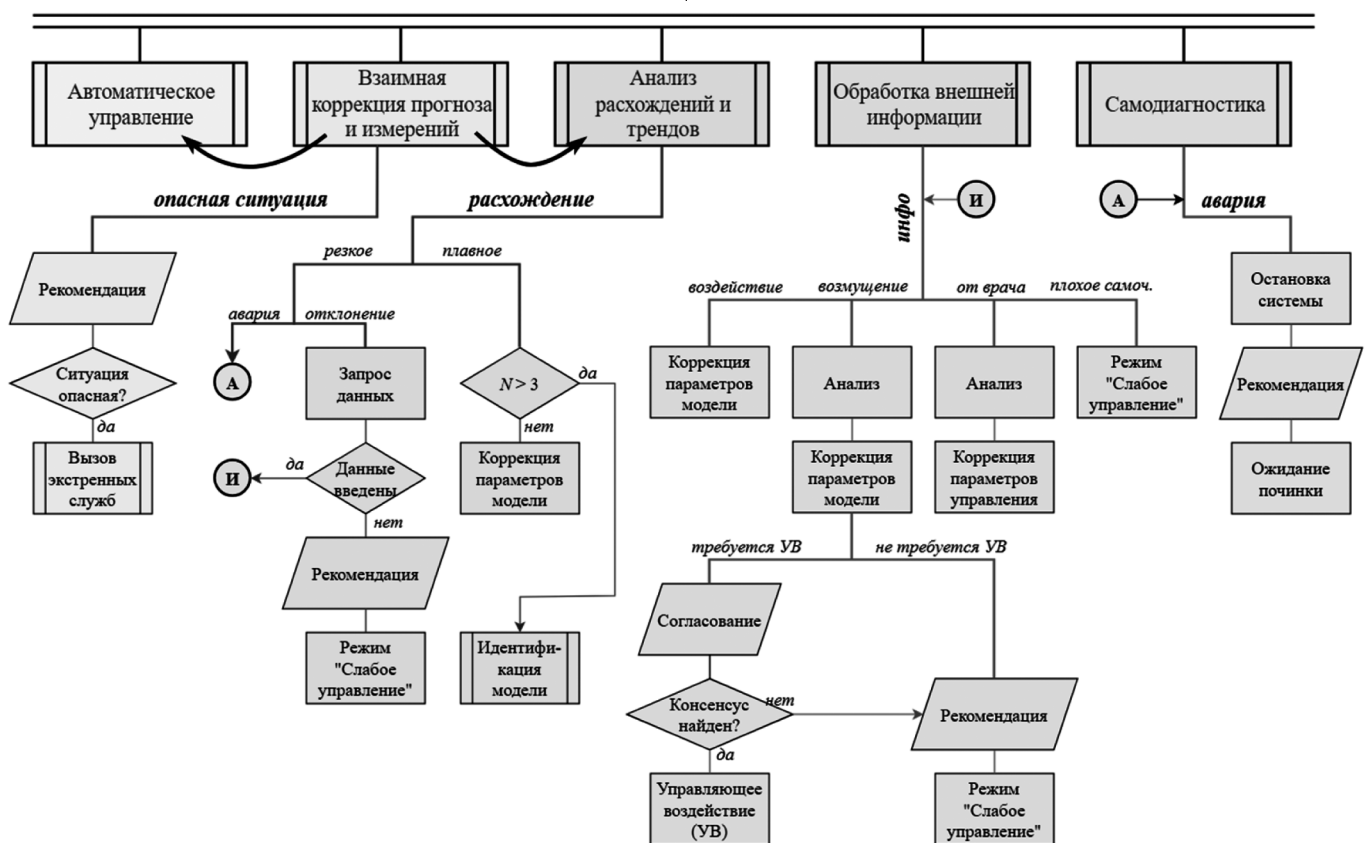


Рис. 1. Алгоритм работы системы управления концентрацией глюкозы в крови и поддержки принятия решений

нам затруднено (сон, стресс, плохое самочувствие и др.), система должна осуществлять управление с большей осторожностью. Таким образом, принцип человеко-машинного консенсуса хотя и имеет потенциал к повышению эффективности управления и снижения нагрузки на пользователя, требует адаптивности системы управления с учётом состояния пользователя.

С точки зрения теории принятия решений система поддержки принятия решений должна отвечать аксиомам теории ожидаемой полезности фон Неймана-Моргенштерна. Ожидаемая полезность должна быть максимизирована, что для рассматриваемой системы на основе оптимизации управляющих воздействий прямо выполняется. Поскольку система при расчётах не учитывает пожелания пользователя ввиду того, что полезность определяется экспертом, а не пользователем, причём однозначным образом, вопрос транзитивности и ранжирования альтернатив не является для данной задачи актуальным, как и вопрос вероятностных оценок.

Существенными для конечной реализации системы на основе предложенного принципа является интерфейс взаимодействия, качество компромиссов, насколько прозрачно, ненавязчиво и обоснованно система предлагает информацию об альтернативах и их послед-

ствиях, помогает ли система корректно расставлять поведенческие приоритеты в пользу повышения общей эффективности управления, может ли система корректно учитывать индивидуальные поведенческие шаблоны пациента, а также общая когнитивная и временная нагрузка на пользователя.

В то время как временная нагрузка может быть оценена с использованием, например метода TLM [8], позволяющего оценить время выполнения задачи в интерфейсе, оценка большинства данных параметров является нетривиальной и плохо формализуемой задачей, которая при этом сохраняет высокую важность при разработке медицинских изделий нового поколения. Основным способом получения информации для обеспечения оценки эффективности являются рандомизированные контролируемые исследования, лонгитюдные когортные исследования, методы A/B тестирования, интервьюирование, сбор текстовой обратной связи от пользователей. Стандартом в области рассматриваемых систем являются валидированные опросники, например NASA-TLX [9] для оценки когнитивной нагрузки, опросники удобства использования систем на основе шкалы Ликерта [10, 11]. Для оценки точности предлагаемых решений как правило проводится сопоставление рекомендаций системы с решениями экспертов. Тем не менее актуальной задачей является разработка методов оценки эффектив-

ности, не использующих статистику по пациентам после внедрения. В контексте внедрения принципа человеко-машинного консенсуса это позволит проводить оптимизацию диалога на этапе разработки систем, а не их подстройку после внедрения.

Другими актуальными вопросами для дальнейшего развития предлагаемого принципа в контексте медицинских систем управления является возможность раздельной оценки влияния технической системы и системы поддержки принятия решений на эффективность терапии, вопросы ограниченной рациональности [12, 13], связанные с ограниченными когнитивными способностями пользователя в том числе ввиду нарушений работы организма, а также вопросы доверия к технической системе.

Заключение

Принцип человеко-машинного консенсуса предлагает новый взгляд на создание систем поддержки принятия решений при управлении персонализированными медицинскими системами, в том числе введения инсулина. Ограниченная компетентность пользователей в понимании процессов с сложных биологических системах и в управлении ими затрудняет внедрение высокоэффективных терапевтических систем, использующих нелинейные модели. Предложенный принцип позволяет распределять ответственность между пользователем и технической системой, обеспечивая переход от иерархической структуры их взаимодействия к синергетической, что открывает перспективы для создания и внедрения медицинских систем нового поколения.

ЛИТЕРАТУРА

1. К.В. Пожар. Проблемы и методы автоматизации систем управления концентрацией глюкозы в крови // Приборы и системы. Управление, контроль, диагностика. 2025. №12. С. 11–18.
2. Strukova É.I., Pozhar K.V. A structural model of glucose regulation for building prognostic algorithms for controlling insulin therapy // Biomedical Engineering. 2024. Vol. 58. №. 1. P. 62–67.
3. К.В. Пожар., Д.А. Чупраков. Методы оптимизации вычислений для автоматизации управления инсулинотерапией // Computational nanotechnology. 2025. №2. С. 48–57.
4. Литинская Е.Л. Разработка системы персонализированной инсулинотерапии // Известия высших учебных заведений. Электроника. 2021. № 2 (26). С. 162–171.
5. Литинская Е.Л., Козлов М.Ю. Система поддержки принятия решений по контролю и компенсации динамики концентрации глюкозы в крови // Перспективы науки. 2020. № 12 (135). С. 54–58.
6. А.П. Коняева, Э.И. Струкова, К.В. Пожар. Анализ влияния внешних и внутренних воздействий на концентрацию глюкозы в крови у пациентов с сахарным диабетом в системах автоматизированной инсулинотерапии // Перспективы науки. 2025. Т.287. № 4. С.57–63.
7. Rustad J.K., Musselman D.L., Nemeroff C.B. Decision-making in diabetes mellitus type 1 // The Journal of Neuropsychiatry and Clinical Neurosciences. 2013. Vol. 25, № 1. P. 40–50.
8. Rice A.D., Lartigue J.W. Touch-level model (TLM) evolving KLM-GOMS for touchscreen and mobile devices // Proceedings of the 2014 ACM Southeast Conference. 2014. P. 1–6.
9. Hart S.G., Staveland L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research // Advances in Psychology. — North-Holland, 1988. Vol. 52. P. 139–183.
10. Brooke J., Jordan P.W., Thomas B., et al. SUS-A quick and dirty usability scale / J. Brooke [et al.] // Usability Evaluation in Industry. 1996. P. 4–7.
11. Bangor A., Kortum P.T., Miller J.T. An empirical evaluation of the system usability scale // International Journal of Human–Computer Interaction. 2008. Vol. 24, № 6. P. 574–594.
12. Bonilauri Ferreira A.P.R. et al. Clinical reasoning in the real world is mediated by bounded rationality: implications for diagnostic clinical practice guidelines / A.P.R. Bonilauri Ferreira [et al.] // PloS one. 2010. Vol. 5, № 4. P. e10265.
13. Thomas E.C., Bass S.B., Siminoff L.A. Beyond rationality: expanding the practice of shared decision making in modern medicine // Social Science & Medicine. 2021. Vol. 277. P. 113900.

© Пожар Кирилл Витольдович (pozhar@bms.zone)

Журнал «Современная наука: актуальные проблемы теории и практики»

ПОМЕХОУСТОЙЧИВОСТЬ OFDM-МОДУЛЯЦИИ ПРИ ПЕРЕДАЧЕ ВИДЕОСИГНАЛА В УСЛОВИЯХ РЕТРАНСЛИРОВАННЫХ ПОМЕХ: ИМИТАЦИОННОЕ ИССЛЕДОВАНИЕ В GNU RADIO

Поташин Александр Александрович

Специалист, РТУ МИРЭА

apotashin@list.ru

NOISE-RESISTANT OFDM-MODULATION OF VIDEO SIGNAL TRANSMISSION UNDER CONDITIONS OF RETRANSLATED NOISE: SIMULATION STUDY IN GNU RADIO

A. Potashin

Summary. The article investigates the noise immunity of OFDM modulation in the transmission of a video signal in the conditions of retransmitted interference. Based on the simulation model implemented in the GNU Radio environment, a comparative measurement of the bit error rate (BER) and the signal-to-noise ratio (SNR) in the range from – 15 to 30 dB — without interference and with the addition of three delayed copies of the signal with amplitude coefficients of 0.05; 0.07 and 0.09 is carried out. It has been established that retransmitted interference systematically increases BER at all SNR values, but its effect is mitigated at $\text{SNR} \geq 25$ dB. To achieve $\text{BER} \leq 0.01$ in an interference-affected channel, the required SNR is 5–7 dB higher than in a channel without interference.

Keywords: OFDM, BER, SNR, retransmitted interference, GNU Radio, cyclic prefix, constellation diagram, multipath propagation, DVB-T, 8-PSK.

Аннотация. В статье исследуется помехоустойчивость OFDM-модуляции при передаче видеосигнала в условиях ретранслированных помех. На основе имитационной модели, реализованной в среде GNU Radio, проведено сравнительное измерение вероятности битовой ошибки (BER) и отношения сигнал/шум (SNR) в диапазоне от – 15 до 30 дБ — без помех и при добавлении трёх задержанных копий сигнала с амплитудными коэффициентами 0,05; 0,07 и 0,09. Установлено, что ретранслированные помехи систематически увеличивают BER при всех значениях SNR, однако при $\text{SNR} \geq 25$ дБ их влияние нивелируется. Для достижения $\text{BER} \leq 0,01$ в канале с помехами требуется SNR на 5–7 дБ выше, чем без них.

Ключевые слова: OFDM, BER, SNR, ретранслированные помехи, GNU Radio, циклический префикс, диаграмма созвездия, многолучевое распространение, DVB-T, 8-PSK.

Введение

Технология OFDM получила широкое распространение в современных системах цифровой передачи данных: она лежит в основе стандартов цифрового телевидения DVB-T/T2, беспроводных сетей Wi-Fi (IEEE 802.11a/g/n/ac/ax), стандартов мобильной связи LTE и 5G NR [1, с. 84; 2, с. 70]. Ключевыми достоинствами OFDM являются высокая спектральная эффективность, достигаемая за счёт ортогонального перекрытия поднесущих, устойчивость к межсимвольной интерференции (ISI) благодаря введению циклического префикса, а также возможность простого одноотводного выравнивания канала в частотной области [1, с. 83; 3, с. 1558].

Вместе с тем в условиях плотной городской застройки сигнал, передаваемый с использованием OFDM, подвергается воздействию ретранслированных помех — задержанных и ослабленных копий исходного сигнала, формирующихся при его отражении от зданий, рельефа местности и других объектов. При превышении задержкой отражённой копии длительности циклического префикса ортогональность поднесущих нарушается, воз-

никает межканальная интерференция (ICI), снижается эффективное отношение сигнал/шум и резко возрастает вероятность битовой ошибки [4, с. 35]. Это непосредственно сказывается на качестве видеосигнала: появляются артефакты изображения, блокинг-эффекты, рассинхронизация аудио и видео.

В научной литературе накоплен значительный массив исследований по OFDM: пространственному мультиплексированию в системах MIMO-OFDM [5, с. 686], алгоритмам выравнивания канала [6], методам снижения PAPR [7, с. 84]. Однако работ, систематически исследующих влияние ретранслированных помех на BER при передаче видеосигнала с использованием реалистичной имитационной модели, явно недостаточно [4, с. 40]. В большинстве теоретических работ помехи рассматриваются в рамках упрощённых аналитических моделей, тогда как поведение системы в условиях составной помехи (AWGN + несколько задержанных копий) требует численного моделирования.

В настоящей работе предложена и верифицирована имитационная модель, реализованная в среде GNU

Radio, которая позволяет воспроизвести передачу реального видеофайла через канал с ретранслированными помехами и количественно оценить деградацию BER. Модель охватывает полный тракт: формирование OFDM-символов, введение трёх задержанных копий, добавление AWGN-шума, демодуляцию, измерение BER и SNR.

Цель работы: количественное исследование влияния ретранслированных помех на BER OFDM-системы при передаче видеосигнала в диапазоне SNR от 15 до 30 дБ, а также разработка и верификация соответствующей имитационной модели.

Теоретические основы OFDM и ретранслированных помех

1. Математическая модель OFDM-сигнала

Принцип OFDM состоит в разбиении широкополосного потока данных на N параллельных подпотоков, каждый из которых модулируется на собственной поднесущей частоте. Передаваемый OFDM-сигнал в k -м символьном интервале [1, с. 90]:

$$\dot{u}(t) = \sum_{i=0}^{N_s-1} d_i \exp\left\{j2\pi \frac{i}{T}(t - t_k)\right\} \quad (1)$$

где d_i — комплексный символ на i -й поднесущей, T — полезная длительность символа. Взаимная ортогональность поднесущих обеспечивается выбором $\Delta f = 1/T$. В частотной области каждая поднесущая имеет форму sinc-функции, обнуляющейся на частотах всех соседних поднесущих — как показано на рисунке 1.

Для устранения межсимвольной интерференции к каждому символу добавляется циклический префикс (CP) длительностью $T_g = D \times T_c$. Коэффициент полезного использования мощности [1, с. 88]:

$$\eta_D = \frac{T_u}{T_s} = \frac{N_{FFT}}{(N_{FFT} + D)} \quad (2)$$

В данной работе $N_{FFT} = 128$, $D = 32$, откуда $\eta_D = 0,80$. Принятый сигнал на выходе БПФ для n -го подканала [1, с. 86; 3, с. 1561]:

$$Z_n = H_n \times w_n \times A_n + x_n \quad (3)$$

где H_n — коэффициент передачи канала, w_n — весовой коэффициент, A_n — переданный символ, x_n — аддитивный шум. При $\tau_i > T_g$ в выражении (3) появляется дополнительная составляющая ICI.

2. Математическая модель ретранслированной помехи

Гармоническая помеха с относительной интенсивностью μ , частотной расстройкой $\Delta\omega_{int}$ и случайной начальной фазой φ_{int} [4, с. 39]:

$$s_{int} = \mu A_0 \cos[(\omega_0 + \Delta\omega_{int})t + \varphi_{int}] \quad (4)$$

Ретранслированная помеха — задержанная копия передаваемого сигнала [4, с. 37]:

$$s_{int}(t) = \mu S(t - \tau, \varphi_{int}) \quad (5)$$

где μ — интенсивность, τ — задержка, φ_{int} — начальная фаза. Суммарный принятый сигнал при M задержанных копиях:

Спектр OFDM-сигнала: ортогональные поднесущие (sinc-функции)

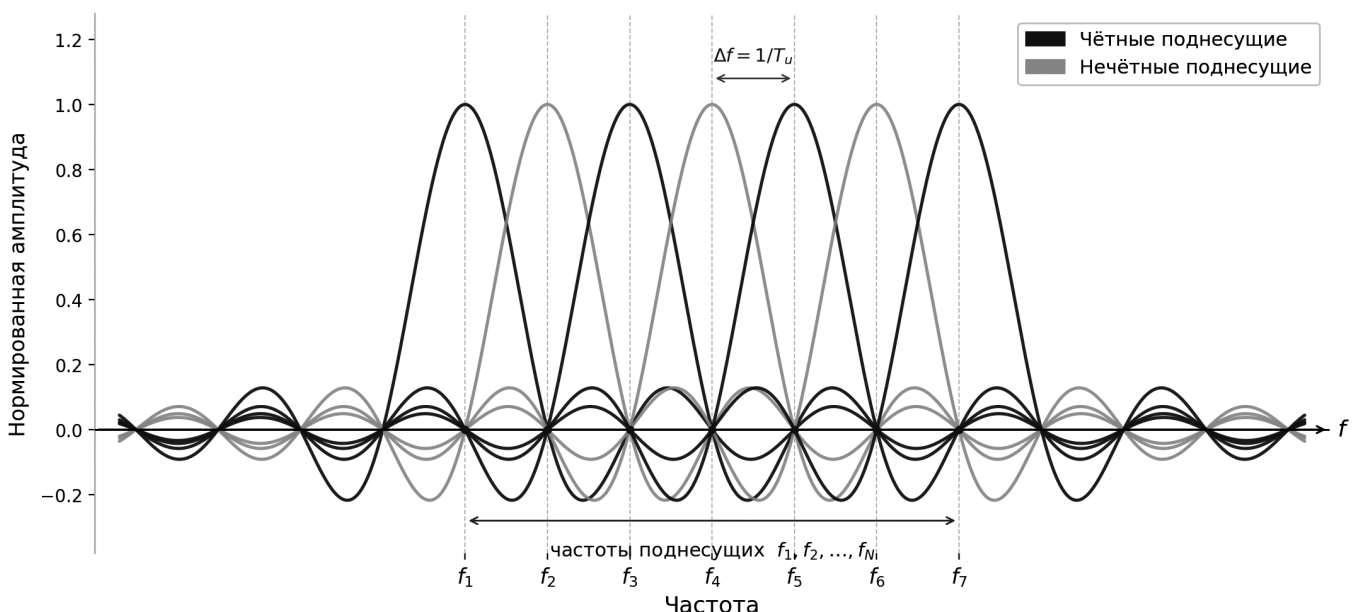


Рис. 1. Спектр OFDM-сигнала: ортогональные поднесущие (sinc-функции); $\Delta f = 1/T$ — межподнесущий интервал

$$r(t) = s(t) + \sum_{i=1}^M \mu_i S(t - \tau_i) + n(t) \quad (6)$$

где $n(t)$ — AWGN. Суммарная мощность помех в данной модели ($M = 3$, $\mu_1 = 0,05$, $\mu_2 = 0,07$, $\mu_3 = 0,09$):

$$P_{int} = \sum_{i=1}^M \mu_i^2 \times P_s \approx 0,0155 \times P_s \quad (7)$$

около 1,6 % от мощности полезного сигнала — достаточно для заметной деградации BER при низком SNR.

3. Вероятность битовой ошибки для 8-PSK

Для схемы 8-PSK теоретическая BER в канале AWGN:

$$BER \approx \left(\frac{2}{\log_2 M} \right) \times Q \left(\sqrt{\frac{2 \log_2 M \times E_b}{N_0}} \right) \times \sin \left(\frac{\pi}{M} \right) \quad (8)$$

где $Q(x)$ — функция Гаусса. При $E_b/N_0 = 10$ дБ теоретическое $BER \approx 0,028$; при 15 дБ — около 0,001. Экспериментальные значения (таблица 2) при $SNR = 12,5$ дБ: без помех — 0,030, с помехами — 0,043, что согласуется с формулой (8).

4. Сравнение OFDM с одноносительными схемами

По данным источника [8, с. 7], при восьмиотводном многолучевом канале BER одноносительных схем составил: BPSK — 0,0072, QPSK — 0,1435, 16QAM — 0,3241, 64QAM — 0,3762 (таблица 1). OFDM превосходит эти схемы, преобразуя широкополосный частотно-избирательный канал в набор плоскостазмирующих узкополосных подканалов, каждый из которых эквализируется делением на одно комплексное число N.

Таблица 1.

Сравнение BER одноносительных схем при многолучевом канале (8 отводов) [8, с. 6]

Схема	BER	Спектр. эффективность, бит/с/Гц	Устойчивость к многолучевости
BPSK	0,0072	1	Высокая
QPSK	0,1435	2	Средняя
16QAM	0,3241	4	Низкая
64QAM	0,3762	6	Очень низкая

1. Общая архитектура

Для имитационного моделирования выбрана платформа GNU Radio — среда разработки программно определяемого радио, реализующая потоковую обработку сигналов в реальном времени. Параметры модели: samp_rate = 100 кГц; NFFT = 128; D = 32 ($T_g = 320$ мкс); модуляция заголовка — BPSK; данных — 8-PSK. Полная схема представлена на рисунке 2.

Источником данных служит видеофайл MP4 (File Source, длина пакета 128 байт). Поток преобразуется в формат с пакетными тегами (Stream to Tagged Stream, длина 50) и поступает на OFDM Transmitter. Исходный байтовый поток записывается в виртуальный буфер «ishod» для последующего сравнения.

2. Модель ретранслированных помех

Ретранслированные помехи формируются тремя параллельными ветвями: Multiply Const ($\mu_1 = 0,05$; $\mu_2 = 0,07$; $\mu_3 = 0,09$) и Delay ($\tau_1 = 3$, $\tau_1 = 2$, $\tau_1 = 1$ отсчёта). Три копии суммируются через Add и поступают на Channel Model с AWGN (рисунок 3).

3. Измерение BER и SNR

Схема измерения BER (рисунок 4) сравнивает сигналы «ishod» и «preobr» через блок BER; результат отображается на QT GUI Number Sink. Схема измерения SNR (рисунок 5) использует MPSK SNR Estimator (тип Simple, 10 000 отсчётов между тегами).

Верификация имитационной модели

До проведения основного эксперимента выполнена верификация корректности измерительной цепи. Из схемы исключены ветви ретранслированных помех (рисунок 6): при нулевом уровне шума разность исходного и демодулированного сигналов должна тождественно равняться нулю.

Результаты верификации (рисунок 7): верхний осциллограф фиксирует нулевую амплитуду разности на всём интервале 0 — 10 мс, значение $BER = 0,000$. Верификация пройдена успешно.

Результаты исследования

1. Спектр принятого сигнала и диаграмма рассеяния

Рисунок 8 демонстрирует спектр мощности принятого OFDM-сигнала (QT GUI Frequency Sink, FFT = 1024). Спектр имеет прямоугольную огибающую в полосе от -3 до +3 МГц, уровень около -100...-103 дБВт/Гц. За пределами полосы спектральная плотность снижается на 20–25 дБ, что подтверждает корректную работу OFDM Transmitter.

Диаграмма рассеяния (рисунок 9) при $SNR \approx 5$ –8 дБ с ретранслированными помехами: синие точки — принятые символы, красные — идеальные позиции 8-PSK. Отсутствие различных кластеров свидетельствует о деградации сигнала — шум и фазовые искажения коэффициентов N размывают созвездие, что соответствует высоким значениям $BER \approx 0,06$ — 0,12 в этой зоне SNR.

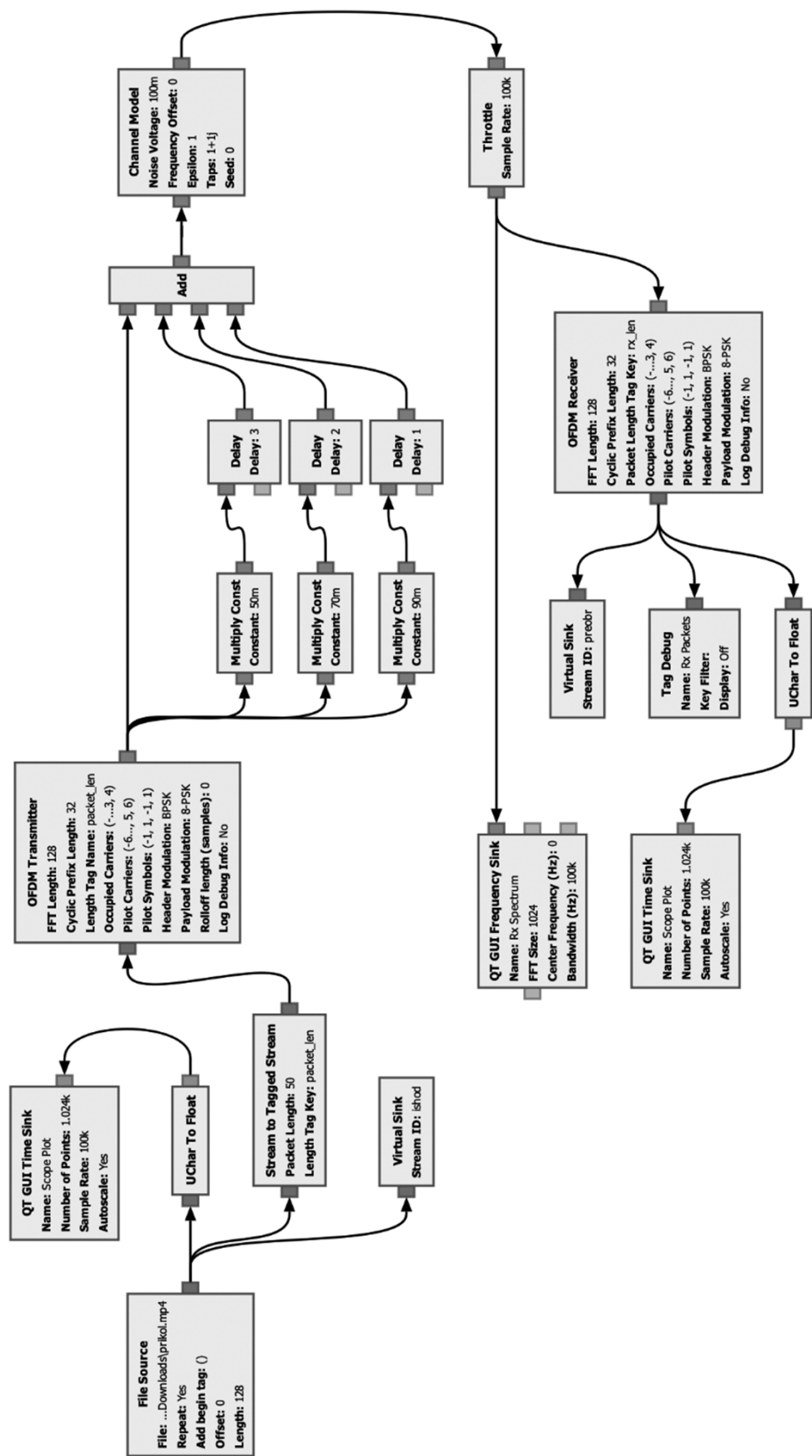


Рис. 2. Имитационная схема в GNU Radio: полный тракт передачи OFDM-видеосигнала с ретранслированными помехами

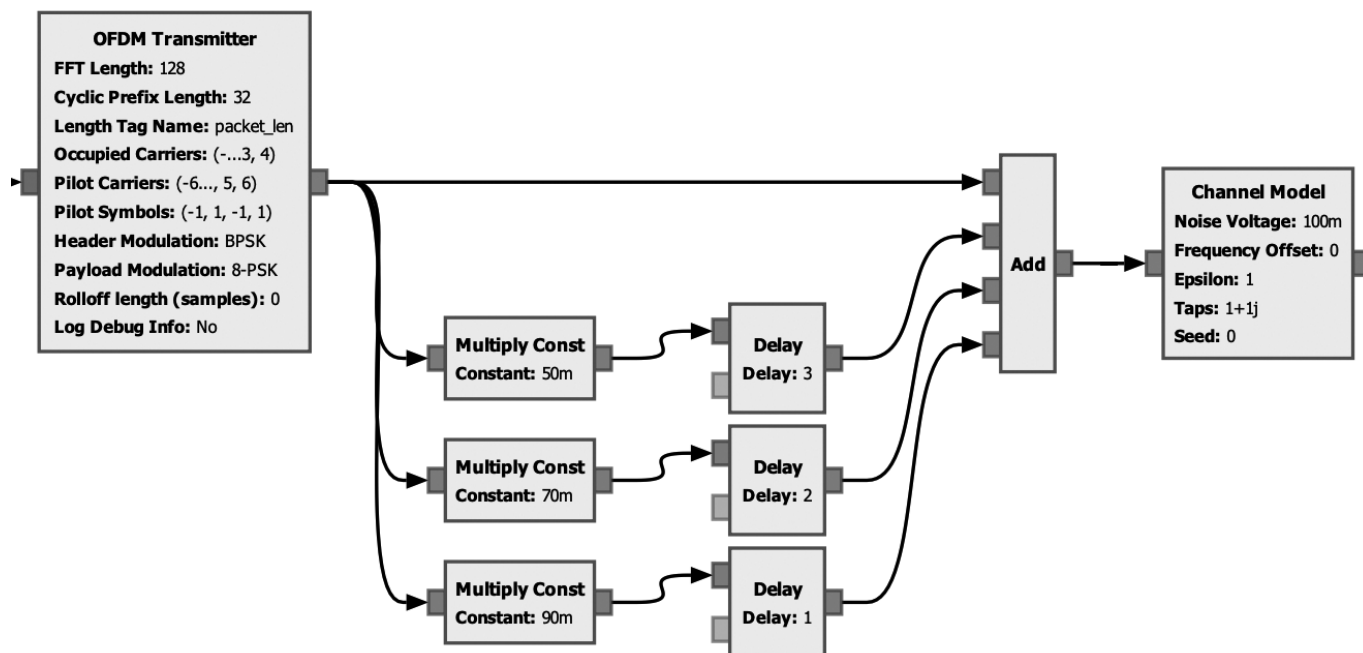


Рис. 3. Блок-схема формирования ретранслированных помех

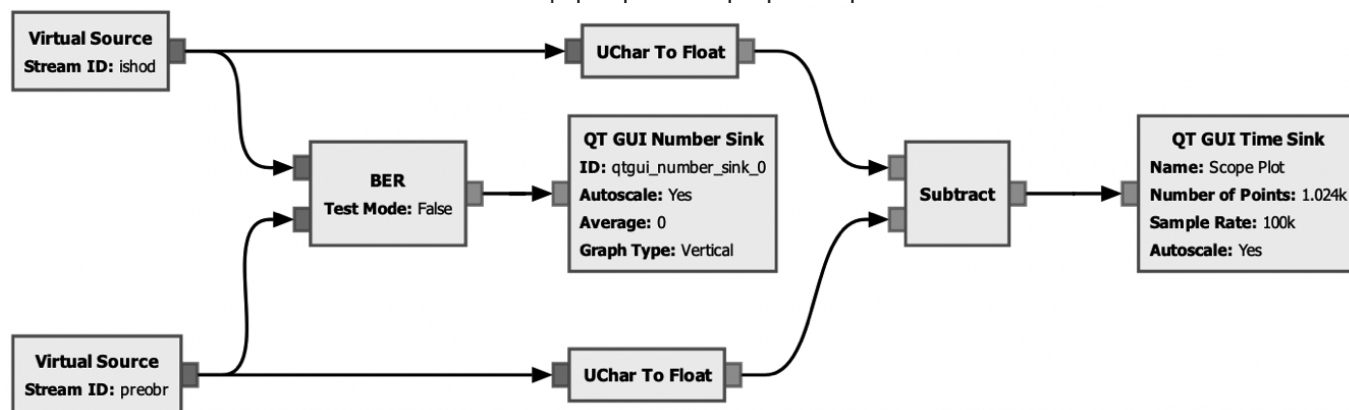


Рис. 4. Схема измерения BER

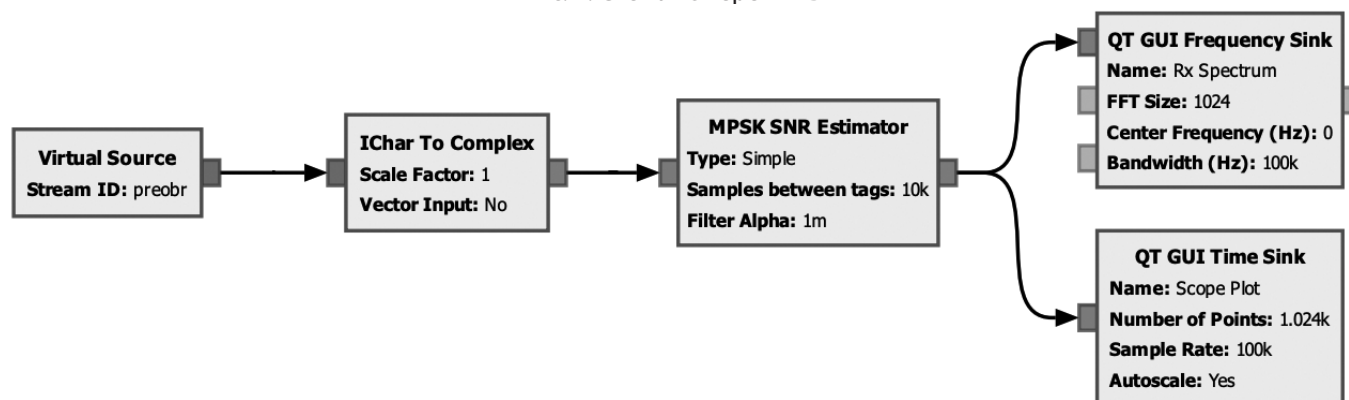


Рис. 5. Схема измерения SNR

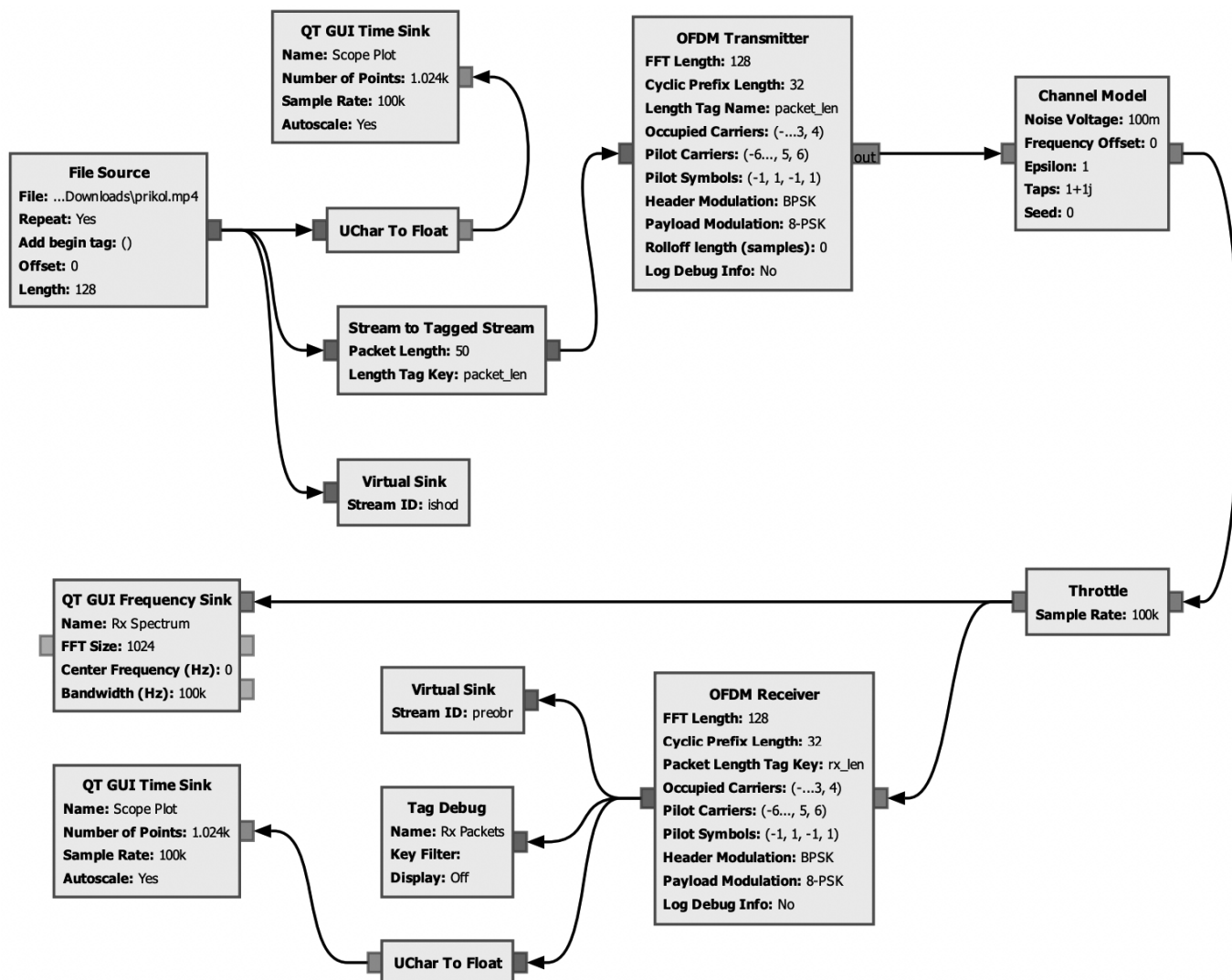
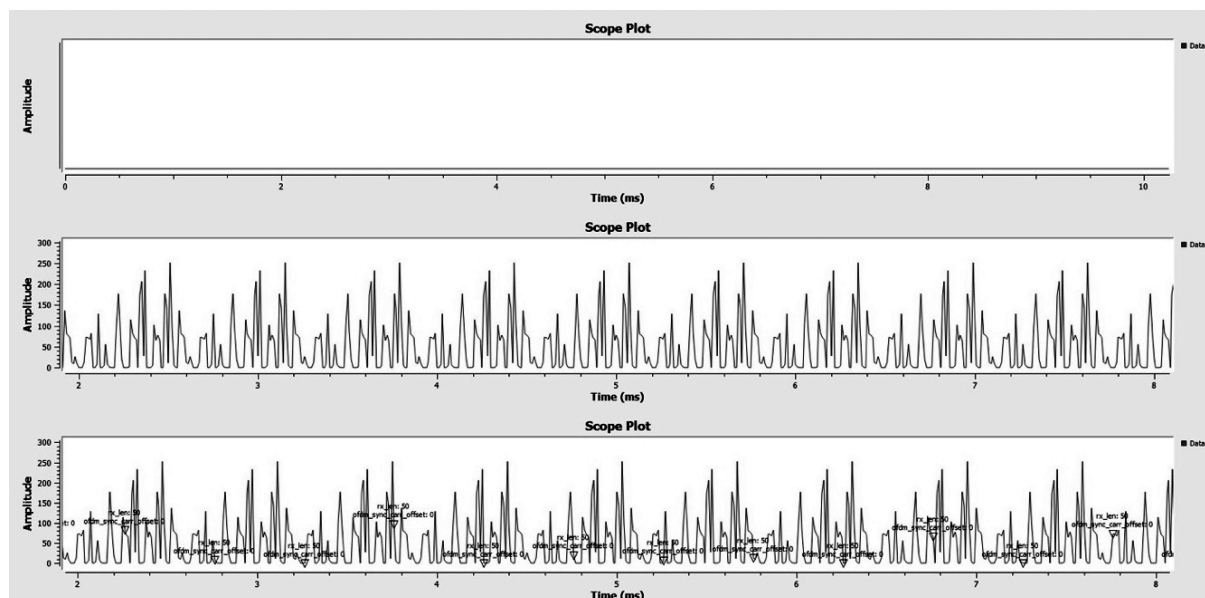


Рис. 6. Верификационная схема без ретранслированных помех

Рис. 7. Результаты верификации: верхний — разность ($\equiv 0$), средний — исходный, нижний — демодулированный без помех

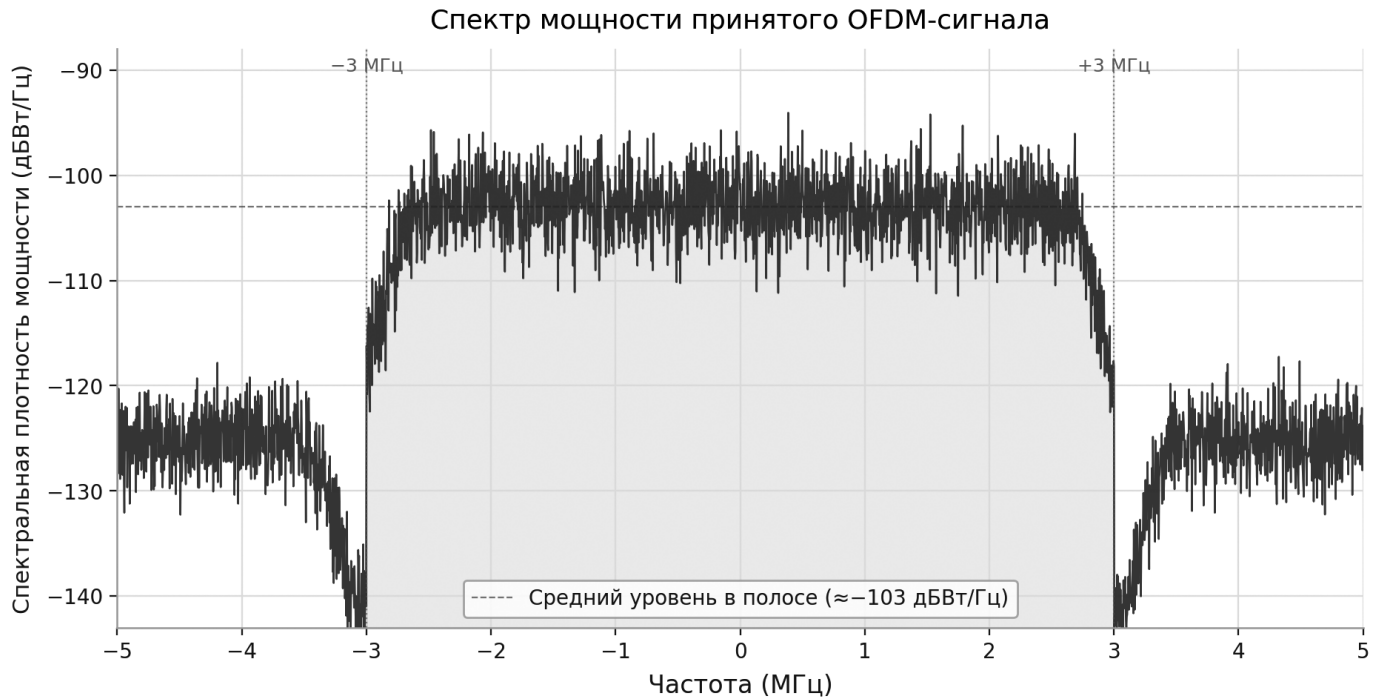
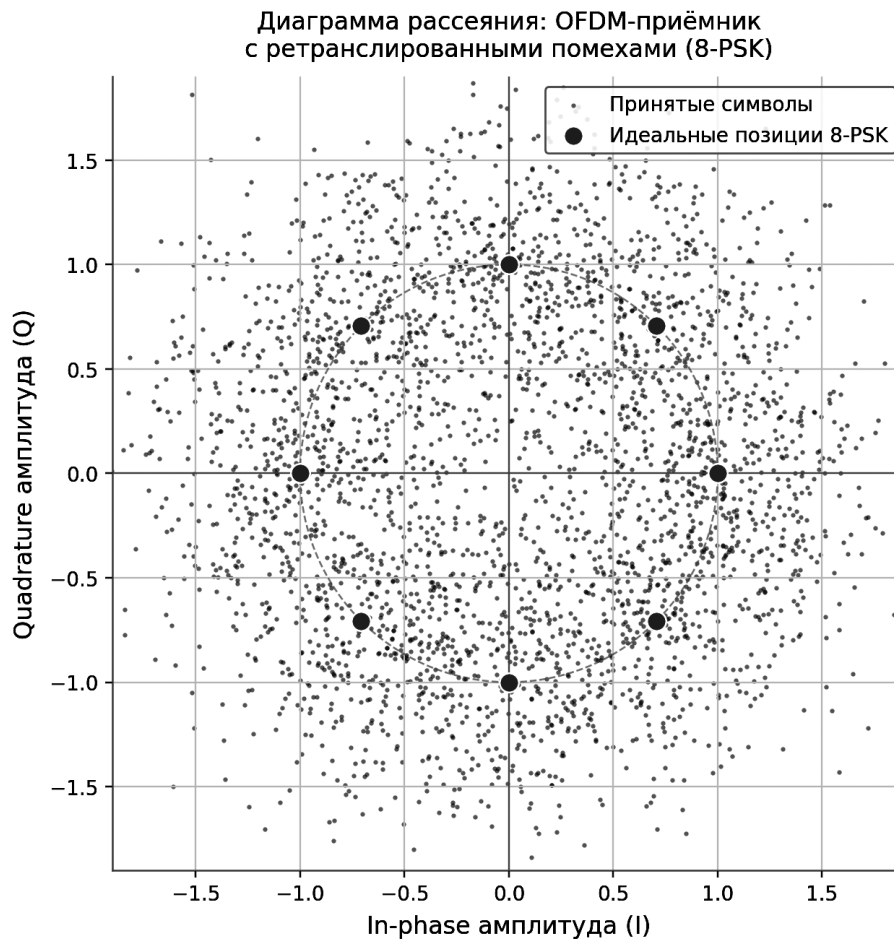


Рис. 8. Спектр мощности принятого OFDM-сигнала (GNU Radio, NFFT = 1024, samp_rate = 100 кГц)

Рис. 9. Диаграмма рассеяния: OFDM-приёмник при $\text{SNR} \approx 5\text{--}8$ дБ с ретранслированными помехами (8-PSK)

2. Зависимость BER от SNR

Результаты 19-точечного измерения BER при варьировании SNR от -15 до 30 дБ представлены в таблице 2. Каждое значение получено при установившемся режиме (не менее 30 с) путём усреднения показаний QT GUI Number Sink.

Таблица 2.
Зависимость BER от SNR при передаче OFDM-видеосигнала (8-PSK, NFFT = 128)

SNR, дБ	BER без помех	BER с помехами
-15	0,358	0,428
$-12,5$	0,349	0,416
-10	0,324	0,404
$-7,5$	0,261	0,388
-5	0,213	0,366
$-2,5$	0,174	0,334
0	0,123	0,270
2,5	0,102	0,200
5	0,080	0,115
7,5	0,060	0,076
10	0,040	0,055
12,5	0,030	0,043
15	0,018	0,024
17,5	0,012	0,017
20	0,009	0,013
22,5	0,006	0,012
25	0,004	0,008
27,5	0,003	0,006
30	0,001	0,001

3. Анализ результатов

При $\text{SNR} < 0$ дБ оба варианта дают $\text{BER} > 0,1$ — неприемлемое качество передачи. В диапазоне $2,5$ – 15 дБ ретранслированные помехи примерно удваивают BER: при $\text{SNR} = 5$ дБ прирост 44% ($0,080$ – $0,115$), при $\text{SNR} = 10$ дБ — 37% ($0,040$ – $0,055$). Для достижения $\text{BER} \leq 0,01$ без помех достаточно $\text{SNR} \approx 12,5$ дБ, с помехами — $\text{SNR} \approx 17,5$ – 20 дБ, разница 5 – 7 дБ. При $\text{SNR} \geq 25$ – 30 дБ оба значения сходятся к $\text{BER} = 0,001$. Результаты согласуются с теоретической кривой по формуле (8) с расхождением 3 – 5 дБ.

Заключение

В работе разработана, верифицирована и применена имитационная модель системы передачи OFDM-видеосигнала с ретранслированными помехами в GNU Radio. Модель реализует полный тракт (NFFT = 128, CP = 32, 8-PSK) и воспроизводит три задержанные копии сигнала с амплитудами $0,05$; $0,07$; $0,09$ (суммарная помеховая мощность $\approx 1,6\%$ от Ps). Верификация подтвердила корректность измерительной цепи.

Ретранслированные помехи систематически увеличивают BER во всём диапазоне SNR: наиболее значимый эффект — в диапазоне 5 – 20 дБ (прирост BER на 37 – 44% , смещение порога $\text{BER} \leq 0,01$ на 5 – 7 дБ). При $\text{SNR} \geq 25$ дБ влияние помех нивелируется. Диаграмма рассеяния и спектр принятого сигнала качественно подтверждают количественные данные таблицы 2.

Практическая значимость: результаты позволяют обоснованно назначать требования к энергетическому запасу при проектировании OFDM-систем видеопередачи в условиях городской многолучевости. Перспективные направления: адаптивное bit loading; алгоритмы компенсации канала (LS, MMSE); варьирование параметров задержанных копий; экспериментальная проверка с SDR-устройствами.

ЛИТЕРАТУРА

1. Makarenko A., Qasim N., Turovsky O. Reducing the impact of interchannel interference on the efficiency of signal transmission in telecommunication systems based on OFDM // Eastern-European Journal of Enterprise Technologies. 2023. Vol. 1, No. 9 (121). P. 82–93. DOI: 10.15587/1729-4061.2023.274501
2. Alhyani N.J., Hamid O.K., Yahia S.A., Ayman M.I. Efficient terrestrial digital video broadcasting receivers based OFDM techniques // Przegląd Elektrotechniczny. 2021. R. 97, No. 11. P. 74–77. DOI: 10.15199/48.2021.11.13
3. Dardari D., Martini M.G., Mazzotti M., Chiani M. Layered video transmission on adaptive OFDM wireless systems // EURASIP Journal on Applied Signal Processing. 2004. No. 10. P. 1557–1567. DOI: 10.1155/S1110865704403102
4. Куликов Г.В., До Чунг Тиен. Эффективность адаптивного фильтра с алгоритмом слежения за огибающей при приёме сигналов с многопозиционной ФМ на фоне нефлуктуационных помех // Российский технологический журнал. 2020. Т. 8, № 5. С. 34–43. DOI: 10.32362/2500-316X-2020-8-5-34-43
5. Morinaga K., Tsutsui H., Miyana Y. An evaluation of wireless video transmission using lossless compression and 8×8 MIMO-OFDM wireless transceiver // Proc. ISCIT 2013. P. 685–690. DOI: 10.1109/ISCIT.2013.6645962
6. Usha S.M., Mahesh H.B. Least square channel estimation for image transmission with OFDM over fading channel // International Journal of Computer Networks & Communications. 2020. Vol. 12, No. 1. DOI: 10.5121/ijcnc.2020.12101
7. Hassan G.M., Mukred M., Gumaei A.H. Modified method of PAPR reduction using clipping and filtering for image transmission with OFDM // Al-Mustansiriyah Journal of Science. 2023. Vol. 34, No. 4. P. 75–86. DOI: 10.23851/mjs.v34i4.1400
8. Waliullah G.M., Yasmin N., Hossain M.A., Abdullah I., Bala D. Analysis the performance of OFDM for different modulation techniques with channels and image transmission // Journal of Electrical Engineering, Electronics, Control and Computer Science. 2023. Vol. 9, No. 34. P. 1–8.

© Поташин Александр Александрович (apotashin@list.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ ВИДЕОАНАЛИТИКИ ДЛЯ ОБНАРУЖЕНИЯ ДЕФЕКТОВ КОНВЕЙЕРНОЙ ЛЕНТЫ ТРАНСПОРТИРОВКИ РУДЫ

DEVELOPMENT OF A VIDEO ANALYTICS INFORMATION SYSTEM FOR DETECTING DEFECTS IN ORE CONVEYOR BELTS

N. Toichkin

Summary. The paper presents the architecture and key solutions of a video analytics information system for automatic defect detection of conveyor belts in the mining industry. The relevance of the development is substantiated due to the insufficient efficiency of traditional visual inspection methods and high economic losses from emergency downtime. A multi-level software architecture is proposed, including edge computing, deep learning for multiclass semantic segmentation, inference optimization using TensorRT, defect tracking with the SORT algorithm, time-series storage in TimescaleDB, and integration with automated process control systems via a REST API. Key features of the system include distinguishing damage types (crack, detachment, rivet), automatic confirmation of repair completion, and accumulation of statistics for predictive maintenance. The developed concept provides a foundation for creating an industrial solution that increases the operational reliability of conveyor transport.

Keywords: video analytics, conveyor belt, defect detection, deep learning, semantic segmentation, EfficientNet, FPN.

Тоичкин Николай Александрович

кандидат технических наук, доцент,
Мурманский арктический университет, Anapimty
toichkin@list.ru

Аннотация. В статье представлена архитектура и ключевые решения информационной системы видеоаналитики для автоматического обнаружения дефектов конвейерных лент в горнорудной промышленности. Обоснована актуальность разработки в связи с недостаточной эффективностью традиционных методов визуального контроля и высокими экономическими потерями от аварийных простоев. Предложена многоуровневая программная архитектура, включающая периферийную обработку (edge computing), глубокое обучение для мультиклассовой семантической сегментации, оптимизацию инференса с помощью TensorRT, трекинг дефектов алгоритмом SORT, хранение временных рядов в TimescaleDB и интеграцию с АСУ ТП через REST API. Ключевыми особенностями системы являются различение типов повреждений (трещина, отслоение, заклепка), автоматическое подтверждение факта ремонта и накопление статистики для предиктивного обслуживания. Разработанная концепция создает основу для создания промышленного решения, повышающего надежность эксплуатации конвейерного транспорта.

Ключевые слова: видеоаналитика, конвейерная лента, обнаружение дефектов, глубокое обучение, семантическая сегментация, EfficientNet, FPN.

Введение

В современных горнодобывающих и перерабатывающих производствах конвейерный транспорт является основным средством перемещения сыпучих материалов. Конвейерные ленты, составляющие до 40–60 % стоимости ленточных конвейеров, работают в условиях интенсивных механических нагрузок, абразивного износа и воздействия внешней среды [1]. Скорость движения ленты может достигать 2–3 м/с, а протяженность конвейерных линий на открытых горных работах — нескольких километров [2]. Выход из строя конвейерной ленты приводит к незапланированным простоям, ремонтным работам и значительным экономическим потерям.

Традиционные методы контроля состояния ленты основаны на периодических визуальных осмотрах оператором, что сопряжено с субъективностью оценки, невозможностью круглосуточного мониторинга и вы-

сокими трудозатратами. Существующие контактные устройства (датчики схода ленты, концевые выключатели) обеспечивают лишь качественную оценку отклонений и подвержены износу при частом срабатывании. Более совершенные методы неразрушающего контроля, такие как магнитная дефектоскопия или рентгеновское излучение, позволяют обнаруживать скрытые дефекты, но требуют сложного и дорогостоящего оборудования [3]. Таким образом, существует потребность в автоматизированной системе непрерывного мониторинга, способной в реальном времени обнаруживать, классифицировать и локализовывать дефекты конвейерной ленты.

В последние годы активно развиваются системы видеоаналитики на основе методов компьютерного зрения и глубокого обучения. Как отмечает Чумаков К.В., применение алгоритмов искусственного интеллекта для анализа видеоданных позволяет в реальном времени отслеживать нештатные ситуации и выявлять аномалии на производственных объектах [4]. В обзорной ра-

боте Blazej R. с соавторами систематизировали методы неразрушающего контроля конвейерных лент, среди которых особое внимание уделяется системам технического зрения, инфракрасной термографии и гибридным подходам с использованием искусственного интеллекта [5]. На рынке присутствуют коммерческие решения, такие как система TOM belt (рентгеновский контроль) [6] и устройство мониторинга на основе LiDAR (УМКЛ) [7]. Однако данные системы либо ориентированы на выявление специфических типов дефектов (например, контроль металлических тросов), либо не обладают возможностью семантической классификации повреждений. Таким образом, разработка специализированной системы видеоаналитики, адаптированной к условиям конкретного производства и интегрируемой в существующую инфраструктуру АСУ ТП, представляет собой актуальную научно-практическую задачу.

Аппаратный стек системы видеоаналитики

Аппаратное обеспечение разрабатываемой системы должно обеспечивать непрерывную видеофиксацию поверхности ленты в условиях запыленности, вибрации и нестабильного освещения, а также высокопроизводительную обработку видеопотока в реальном времени. Основные компоненты аппаратного стека представлены в таблице 1 и на схеме (рис. 1).

На рис. 1 представлена диаграмма развертывания (deployment diagram), иллюстрирующая физическую структуру системы и взаимосвязи между аппаратными компонентами.

Программный стек системы видеоаналитики

Программное обеспечение системы реализует полный цикл обработки видеоданных: от захвата кадров

до визуализации результатов и интеграции с внешними системами. Архитектура программного комплекса включает восемь уровней, представленных на рис. 2.

Программное обеспечение разрабатываемой системы реализует полный цикл обработки видеоданных — от захвата кадров до визуализации результатов и интеграции с внешними системами автоматизации. Архитектура программного комплекса может быть представлена в виде восьми последовательных уровней, каждый из которых решает определенный класс задач (рис. 2). Рассмотрим ключевые особенности реализации каждого уровня.

Уровень 1 (Сбор и разметка данных). Базовым условием построения эффективной модели мультиклассовой сегментации является качественно размеченный набор данных, включающий классы: «фон», «трещина», «отслоение» и «заклепка». Для решения задачи целесообразно применение инструмента CVAT, обеспечивающего многоклассовую разметку, распределенную работу и контроль качества.

Уровень 2 (Предобработка и аугментация). На данном этапе выполняется подготовка данных к обучению: изменение размера, нормализация и преобразование цветовых пространств средствами OpenCV. Ключевую роль играет аугментация с использованием библиотеки Albumentations, позволяющей синхронно преобразовывать изображения и маски. Применяются изменение яркости/контраста, гауссово размытие (имитация запыленности), спекл-шум и аффинные трансформации (вибрация камеры).

Уровень 3 (Фреймворк глубокого обучения). В качестве основного фреймворка рекомендуется PyTorch, обеспе-

Таблица 1.

Аппаратные компоненты системы видеоаналитики

Компонент	Рекомендуемое решение	Характеристики и обоснование выбора
Промышленная камера	Камера с матрицей CMOS, разрешение от 2 МП, частотой кадров не менее 30 fps	Обеспечение съемки быстро движущейся ленты (2,56 м/с) с достаточным разрешением для обнаружения дефектов шириной раскрытия от 1 мм
Объектив	С фиксированным фокусным расстоянием, с защитой от брызг и пыли (IP65/67)	Работа в условиях запыленности и вибрации, защита от загрязнений
Осветительное оборудование	Светодиодные линейные модули с регулируемой яркостью	Стабилизация освещения для снижения влияния теней и бликов от металлических элементов (заклепок)
Промышленный компьютер	На базе NVIDIA GPU (например, Jetson AGX Orin или аналогичный)	Обеспечение инференса нейросетевых моделей в реальном времени, поддержка оптимизации через TensorRT
Энкодер	Инкрементальный энкодер, связанный с приводным барабаном	Точная привязка координат дефектов к длине ленты, синхронизация с видеопотоком
Сетевое оборудование	Промышленный коммутатор Ethernet, оптоволоконные линии (при необходимости)	Передача видеоданных и управляющих сигналов, интеграция с АСУ ТП

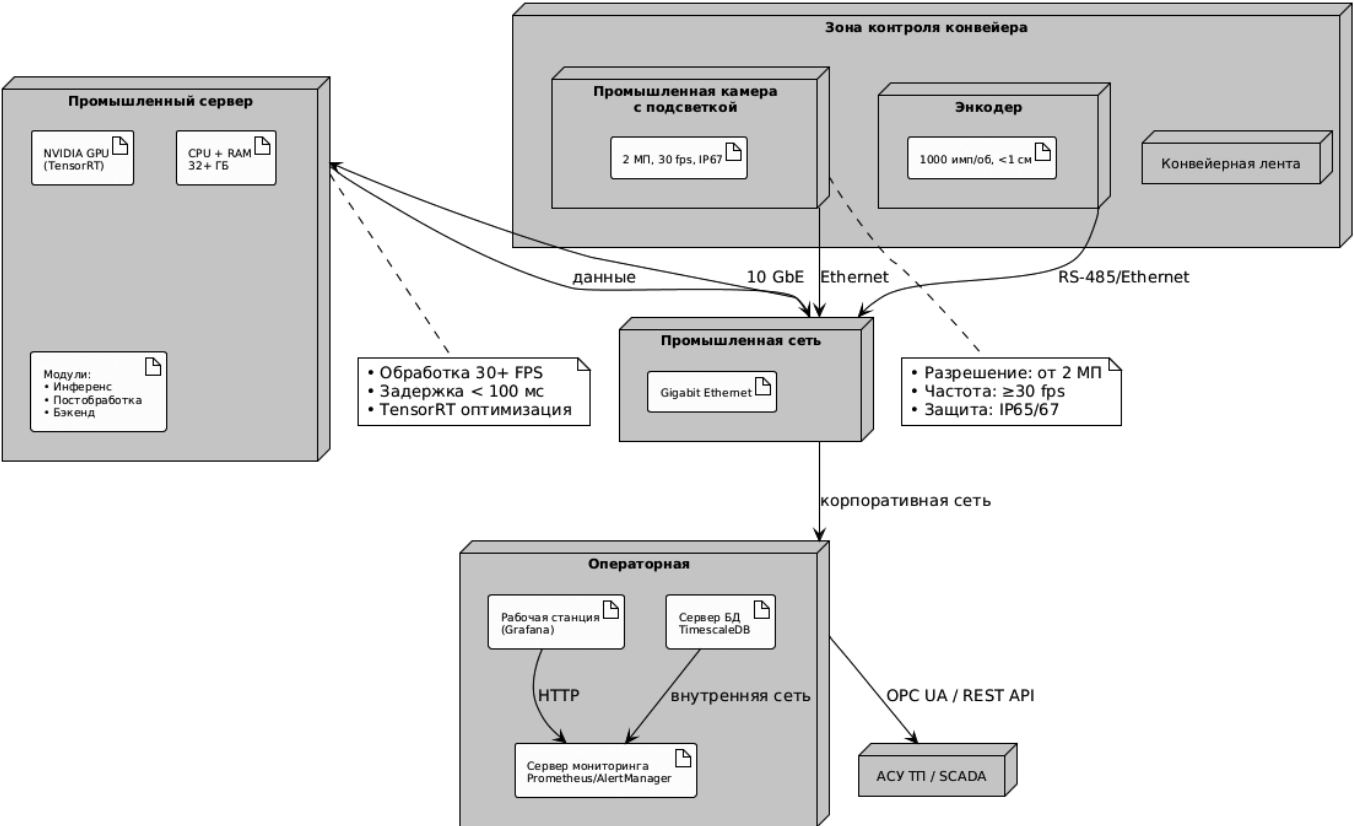


Рис. 1. Схема аппаратного обеспечения системы видеоаналитики

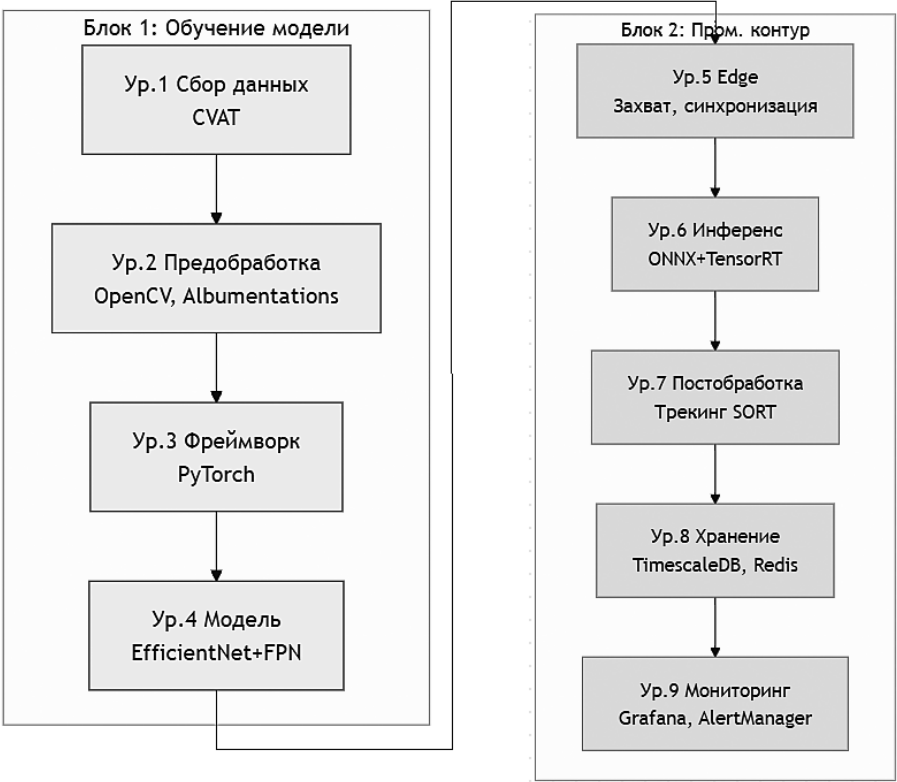


Рис. 2. Архитектура программного обеспечения системы видеоаналитики

чивающий гибкость разработки и поддержку кастомных архитектур. Альтернативно может использоваться TensorFlow при наличии соответствующей экспертизы.

Уровень 4 (Архитектура модели). Для баланса точности и скорости инференса предложена комбинация энкодера EfficientNet-B5 и декодера FPN, обеспечивающая эффективную обработку разномасштабных дефектов. В условиях сильного дисбаланса классов стандартная кросс-энтропия заменена на Focal Loss, фокусирующий обучение на сложных примерах (трещинах и отслоениях).

Уровень 5 (Edge-уровень). Периферийный уровень обеспечивает взаимодействие с аппаратурой и первичную обработку видеопотока в реальном времени. Аппаратная база включает промышленный компьютер с NVIDIA GPU (Jetson AGX Orin), защищенные камеры (IP65/67), светодиодные осветители и инкрементальный энкодер для привязки координат. Функционально уровень решает три задачи: захват видеопотока с частотой ≥ 30 FPS, синхронизацию кадров с энкодером, буферизацию и препроцессинг (изменение разрешения, нормализацию, преобразование цветовых пространств). Результат — подготовленный тензор для нейросетевой модели. Вынесение операций на периферийный уровень снижает требования к пропускной способности сети и обеспечивает детерминированную задержку.

Уровень 6 (Оптимизация и инференс). Для работы в реальном времени модель конвертируется из PyTorch в формат ONNX, обеспечивающий фреймворк-независимость. Финальная оптимизация выполняется с помощью TensorRT, ускоряющего инференс в 2–5 раз, что критически важно для обработки 30 кадров/с.

Уровень 7 (Постобработка и логика). Сырая маска нейросети преобразуется в структурированные данные о дефектах. Средствами OpenCV выполняется поиск контуров для каждого класса. Для разрешения коллизий (например, трещина и заклепка) применяются правила приоритета (заклепка имеет приоритет). Ключевым элементом является трекинг дефектов алгоритмом SORT, позволяющий отслеживать эволюцию повреждений на разных оборотах ленты и автоматически подтверждать факт ремонта (смена класса с «трещины» на «заклепку»).

Уровень 8 (Хранение). Для фиксации типа, координат, размеров и динамики дефектов используется TimescaleDB — расширение PostgreSQL для временных рядов. Кэширование активных дефектов реализовано на Redis. Взаимодействие с внешними системами (SCADA, MES, ERP) обеспечивается через высокопроизводительный REST API на базе FastAPI (рис. 3).

Уровень 9 (Мониторинг и оповещение). Визуализация данных выполняется в Grafana, сбор ме-

трик — Prometheus. Оповещения настраиваются через AlertManager с дифференциацией по классам дефектов: критические сообщения (трещины, отслоения) направляются в мессенджер или SCADA, информация о заклепках только логируется, что предотвращает «усталость оператора» от ложных срабатываний.

Пример кода: базовая реализация модели сегментации

В данном разделе представлен пример реализации модели сегментации дефектов на основе архитектуры U-Net с использованием библиотеки PyTorch. В качестве базовой архитектуры выбрана упрощенная версия U-Net, состоящая из сверточных слоев с активацией ReLU и операций подвыборки (листинг 1).

Листинг 1. Простая U-Net архитектура

```
class SimpleUNet(nn.Module):
    def __init__(self, in_channels=3, out_channels=1):
        super(SimpleUNet, self).__init__()
        self.encoder = nn.Sequential(
            nn.Conv2d(in_channels, 64, kernel_size=3, padding=1),
            nn.ReLU(inplace=True),
            nn.Conv2d(64, 64, kernel_size=3, padding=1),
            nn.ReLU(inplace=True),
            nn.MaxPool2d(kernel_size=2, stride=2)
        )
        self.decoder = nn.Sequential(
            nn.ConvTranspose2d(64, 64, kernel_size=2, stride=2),
            nn.Conv2d(64, 32, kernel_size=3, padding=1),
            nn.ReLU(inplace=True),
            nn.Conv2d(32, out_channels, kernel_size=1)
        )
    def forward(self, x):
        x1 = self.encoder(x)
        x2 = self.decoder(x1)
        return torch.sigmoid(x2)
```

После определения архитектуры модель инициализируется и перемещается на доступное вычислительное устройство (GPU или CPU). В качестве функции потерь используется бинарная кросс-энтропия (BCELoss), оптимизация выполняется методом Adam. Далее реализуется цикл обучения, представленный в листинге 2.

Листинг 2. Цикл обучения

```
for epoch in range(EPOCHS):
    model.train()
    epoch_loss = 0
    for images, masks in train_loader:
        images = images.to(device)
        masks = masks.to(device)
        optimizer.zero_grad()
        outputs = model(images)
        loss = criterion(outputs, masks)
```

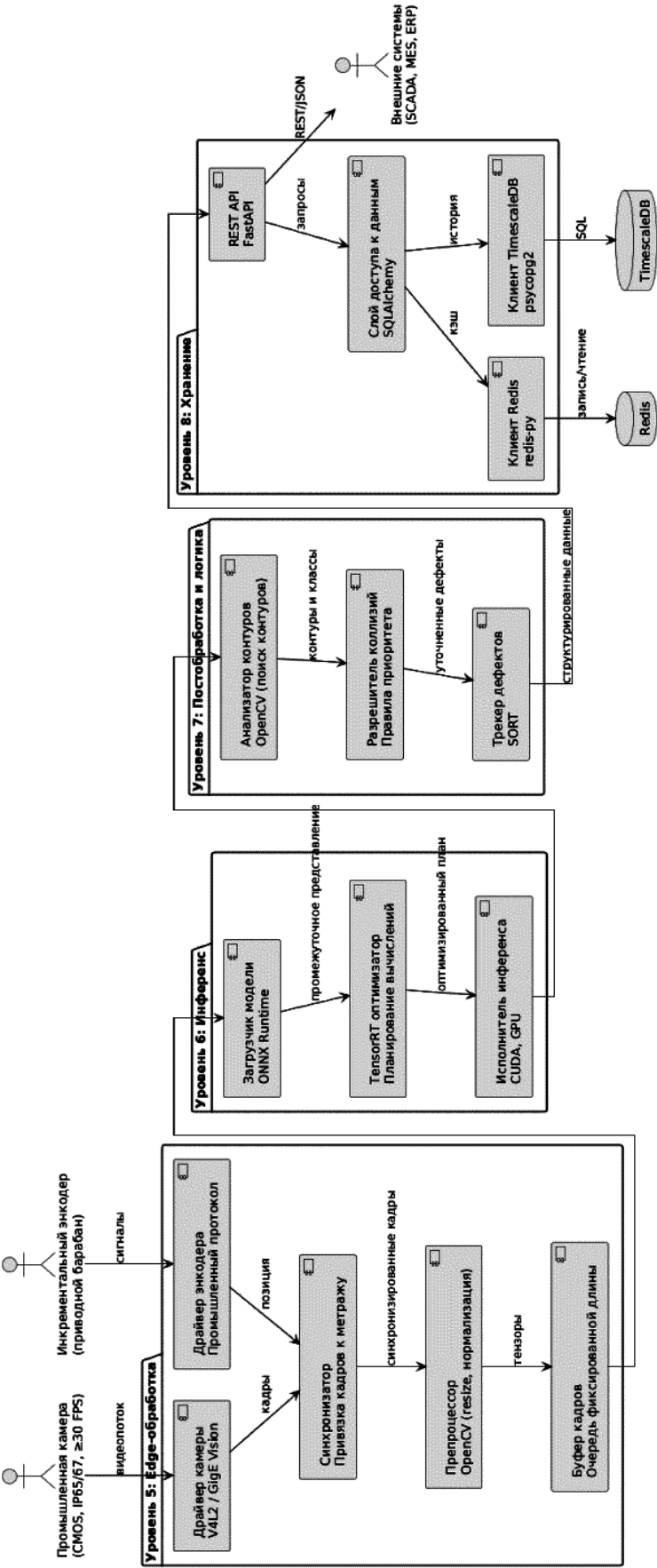



Рис. 3. Диаграмма компонент для блока промышленного контура

```
loss.backward()
optimizer.step()
epoch_loss += loss.item()
print(f'Epoch {epoch+1}/{EPOCHS}, Loss: {epoch_loss/
len(train_loader):.4f}')
```

Для решения задачи мультиклассовой сегментации (трещина, отслоение, заклепка) необходимо модифицировать выходной слой (`out_channels` = число классов), заменить функцию потерь на `CrossEntropyLoss` или `Focal Loss` и использовать маски с целочисленными индексами классов.

Разработанный код может быть масштабирован для промышленного применения с оптимизацией инференса через `ONNX Runtime` или `TensorRT`.

Заключение

В результате исследования разработана архитектура информационной системы видеоаналитики для обнаружения дефектов конвейерных лент. Актуальность

автоматизации подтверждается высокими потерями от простоев и неэффективностью традиционных методов контроля. Аппаратный стек включает промышленные камеры, стабилизированное освещение, компьютер на базе `NVIDIA GPU` и энкодер для привязки координат. Программная архитектура содержит девять уровней: от сбора данных до мониторинга. Ключевые элементы: мультиклассовая сегментация `EfficientNet+FPN` с `Focal Loss`, оптимизация `TensorRT`, `edge`-обработка, трекинг дефектов `SORT` и хранение временных рядов в `TimescaleDB`. Преимущества: различение типов дефектов (трещина, отслоение, заклепка), фильтрация ложных срабатываний, автоматическое подтверждение ремонта, накопление статистики для предиктивного обслуживания и интеграция с АСУ ТП через `REST API`. Дальнейшие исследования: сбор данных, сравнение архитектур нейросетей, калибровка модели и оценка экономической эффективности. Разработанная концепция создает основу для промышленного решения, повышающего надежность конвейерного транспорта и сокращающего затраты на ремонты.

ЛИТЕРАТУРА

1. Zhang M., Jiang K., Zhao S., Hao N., Zhang Y. Deep-learning-based multistate monitoring method of belt conveyor turning section // *Structural Health Monitoring*. — 2024. DOI: 10.1177/14759217231202964
2. Полуниин В.Т., Гуленко Г.Н. Эксплуатация мощных конвейеров. — М.: Недра, 1986. — 240 с.
3. Błażej R., Jurdziak L., Kirjanów-Błażej A. Non-Destructive Testing for Conveyor Belt Monitoring and Diagnostics: A Review // *Applied Sciences*. — 2025. — Vol. 15, No. 24. — Article 13272. DOI: 10.3390/app152413272
4. Чумаков К.В. Тенденции в современных системах мониторинга промышленной безопасности и охраны труда // *Известия Томского политехнического университета. Промышленная кибернетика*. — 2024. — Т. 2, № 4. — С. 34–40. DOI: 10.18799/29495407/2024/4/77
5. Fedorko G., Molnár V., Vasil' M., Salai R. Proposal of digital twin for testing and measuring of transport belts for pipe conveyors within the concept Industry 4.0 // *Measurement*. — 2021. — Vol. 174. — Article 108978. DOI: 10.1016/j.measurement.2021.108978
6. TOM belt Condition Monitoring System // *Phoenix Conveyor Belt Systems*. — URL: <https://www.phoenix-conveyorbelts.com/products/condition-monitoring/tom-belt> (дата обращения: 12.03.2026)
7. Устройство мониторинга конвейерных лент (УМКЛ) // *Северсталь: официальный сайт*. — 30 апреля 2025. — URL: <https://severstal.com/rus/media/archive/karelskiy-okatysh-provel-testirovanie-ustroystva-monitoringa-konveyernykh-lent/> (дата обращения: 12.03.2026)

© Тоичкин Николай Александрович (toichkin@list.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

ИНТЕГРАЛЬНЫЙ ИНДЕКС ЗАГРЯЗНЕНИЯ ПО СПУТНИКОВЫМ ДАННЫМ: АВТОМАТИЗАЦИЯ РАСЧЁТА И ВИЗУАЛИЗАЦИИ ДЛЯ РАССЛЕДОВАНИЯ АВАРИИ НА ХВОСТОХРАНИЛИЩЕ КАТОКА

INTEGRATED POLLUTION INDEX VIA SATELLITE DATA: AUTOMATING CALCULATION AND VISUALIZATION FOR THE INVESTIGATION OF THE CATOCA TAILINGS DAM ACCIDENT

*Tchipungu Eduardo Afonso
V. Shatunova*

Summary. The article presents the results of automated processing of long-term satellite data aimed at assessing the consequences of the accidental breach of the Catoca diamond mine tailings dam (Angola). A methodology for cartographic visualization of the integrated pollution index was developed, ensuring an accurate representation of the spatial distribution of manganese and copper under conditions of significant concentration variability. An analysis of pollution dynamics was performed for three time periods: the baseline period (2002 — July 2021), the acute post-accident period (July 2021 — June 2022), and the long-term period (after June 2022). It is demonstrated that the proposed automation methods provide an objective basis for the environmental investigation of industrial disasters.

Keywords: automated data processing, satellite monitoring, tailings dam, integrated pollution index, maximum permissible concentration (MPC), manganese, copper, accident, Catoca.

Тшипунгу Эдуарду Афонсу

Аспирант, Национальный исследовательский
технологический университет МИСИС, Москва
tchipungo369@mail.ru

Шатунова Вероника Максимовна

Национальный исследовательский технологический
университет МИСИС, Москва

Аннотация. В статье изложены результаты автоматизированной обработки многолетних спутниковых данных, направленной на оценку последствий аварийного разрушения дамбы хвостохранилища алмазодобывающего предприятия Катока (Ангола). Разработана методика картографической визуализации интегрального индекса загрязнения, обеспечивающая корректное отображение пространственного распределения марганца и меди в условиях значительного разброса их концентраций. Выполнен анализ динамики загрязнения за три временных периода: фоновый (2002 — июль 2021), острый послеаварийный (июль 2021 — июнь 2022) и отдалённый (после июня 2022). Показано, что предложенные методы автоматизации создают объективную основу для экологического расследования техногенных катастроф.

Ключевые слова: автоматизация обработки данных, спутниковый мониторинг, хвостохранилище, интегральный индекс загрязнения, предельно допустимая концентрация (ПДК), марганец, медь, авария, Катока.

Введение

Хвостохранилища горнорудных предприятий относятся к объектам повышенной экологической опасности. Аварийные прорывы ограждающих дамб сопровождаются поступлением в водные объекты значительных объёмов загрязняющих веществ, в том числе тяжёлых металлов [1]. Одним из таких объектов служит хвостохранилище алмазодобывающего предприятия Катока в Анголе, где 2 июля 2021 года произошла авария, вызвавшая загрязнение прилегающих водотоков.

Современные методы дистанционного зондирования Земли (ДЗЗ) в сочетании с автоматизированными алгоритмами обработки позволяют оперативно оцени-

вать масштабы загрязнения и отслеживать его пространственно-временную динамику [2]. Особую роль играет картографическая визуализация интегральных показателей, которая даёт возможность наглядно представить распределение поллютантов даже при значительном разбросе концентраций [3]. Цель настоящей работы — разработка и апробация методов автоматизированного анализа и визуализации спутниковых данных для оценки последствий аварии на хвостохранилище Катока.

Материалы и методы исследования

1. Объект исследования

Хвостохранилище предприятия Катока расположено в северо-восточной части Анголы (провинция Южная

Лунда) и является одним из крупнейших в регионе. Авария, произошедшая 2 июля 2021 года, привела к разгерметизации дамбы и сбросу загрязнённых вод в речную сеть. Основными загрязняющими компонентами выступают марганец (Mn) и медь (Cu), содержащиеся в отходах обогащения руды.

2. Исходные данные

В работе использованы спутниковые снимки среднего и высокого разрешения (Landsat-7/8/9, Sentinel-2) за период с 2002 по 2023 год. Обработка включала атмосферную коррекцию, расчёт спектральных индексов и калибровку по данным наземных наблюдений [4]. Для автоматизации вычислений применялись скрипты на языке Python в среде Google Earth Engine, что позволило обработать более 2000 сцен.

3. Периодизация наблюдений

Для анализа динамики весь временной ряд был разделён на три характерных интервала:

- **Период 1 (фоновый):** 1 января 2002 г. — 2 июля 2021 г. (до аварии);
- **Период 2 (острый):** 3 июля 2021 г. — 30 июня 2022 г. (в течение года после аварии);
- **Период 3 (отдалённый):** с 1 июля 2022 г. по декабрь 2023 г. (спустя более года после аварии).

4. Расчёт интегрального индекса загрязнения

Для каждого пикселя вычислялось отношение концентрации вещества к его предельно допустимой концентрации (ПДК) согласно нормативам для водных объектов рыбохозяйственного значения. Интегральный индекс загрязнения определялся как сумма отношений для марганца и меди [5].

5. Методы визуализации

Апробированы два подхода к визуализации:

1. **Линейное и логарифмическое шкалирование** — для отображения всего диапазона значений без потери информации о низких концентрациях [6];
2. **Многослойное наложение** — раздельное построение карт для каждого металла с последующим их совмещением с использованием прозрачности [7].

Результаты и их обсуждение

1. Общая характеристика района исследования

На рис. 1 представлена карта-схема района исследования. Карта выполнена на основе спутниковых сним-

ков высокого разрешения; на ней отмечены границы хвостохранилища, место прорыва дамбы, направление течения реки, а также точки отбора проб, использованные для калибровки спутниковых данных. Данная схема позволяет пространственно локализовать источник загрязнения и основные пути распространения загрязнённых вод.



Рис. 1. Карта-схема района исследования с указанием местоположения хвостохранилища Катока (Ангола) и ключевых створов наблюдения

2. Проблема визуализации концентраций с большим диапазоном значений

При построении среднемноголетних карт концентраций марганца и меди выявлена следующая проблема: из-за значительной разницы в концентрациях этих металлов (марганец может превышать фоновые значения в десятки раз) на линейной шкале области с умеренным загрязнением становятся неразличимы. Это проиллюстрировано на рис. 2 и 3.

На рис. 2 видно, что зоны с высокими концентрациями (приуроченные к месту аварии) занимают доминирующее положение на цветовой шкале, тогда как области с фоновыми и умеренными значениями сливаются в однородный фон. Такое представление не позволяет оценить масштабы распространения загрязнения.

Логарифмическое преобразование (рис. 3) позволяет сжать динамический диапазон данных, благодаря чему на карте одновременно видны как аномально высокие концентрации (отображаемые яркими цветами), так и фоновые значения. Данный подход является предпочтительным для создания обзорных карт интегрального загрязнения, так как он сохраняет информацию о пространственной неоднородности поля концентраций.

3. Сравнение двух вариантов визуализации

Для решения проблемы большого разброса концентраций предложены два подхода. Первый вариант (логарифмическое масштабирование) реализован

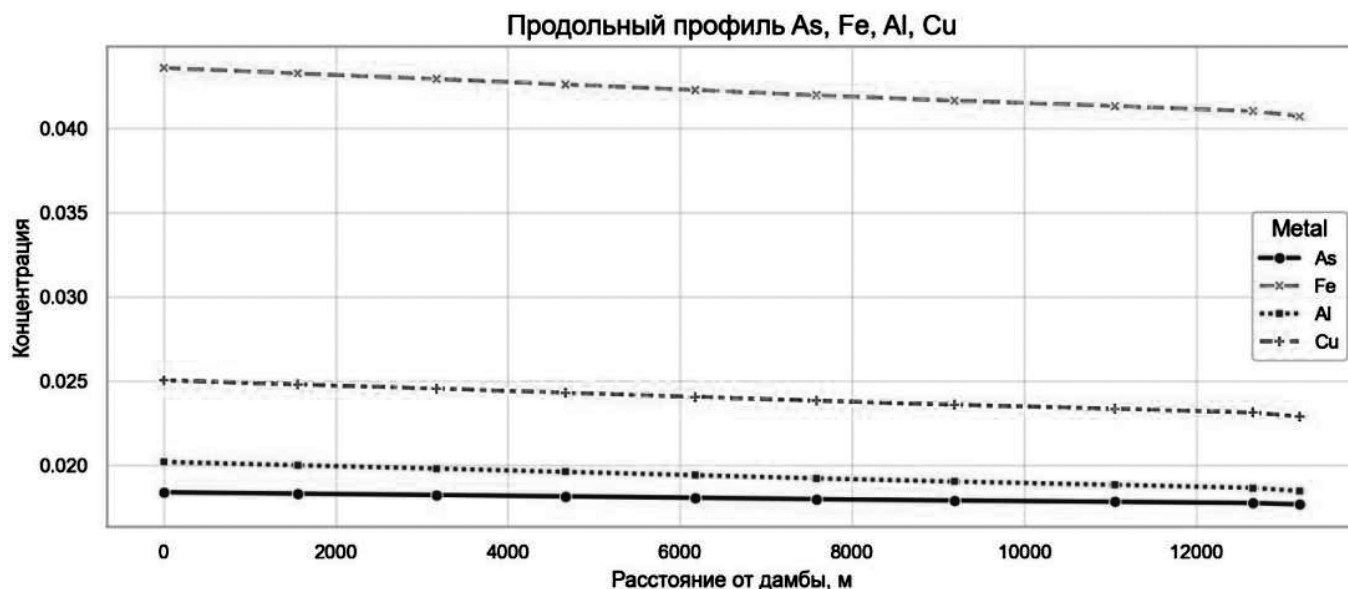


Рис. 2. Пространственное распределение среднееголетних концентраций марганца и меди с использованием линейной шкалы

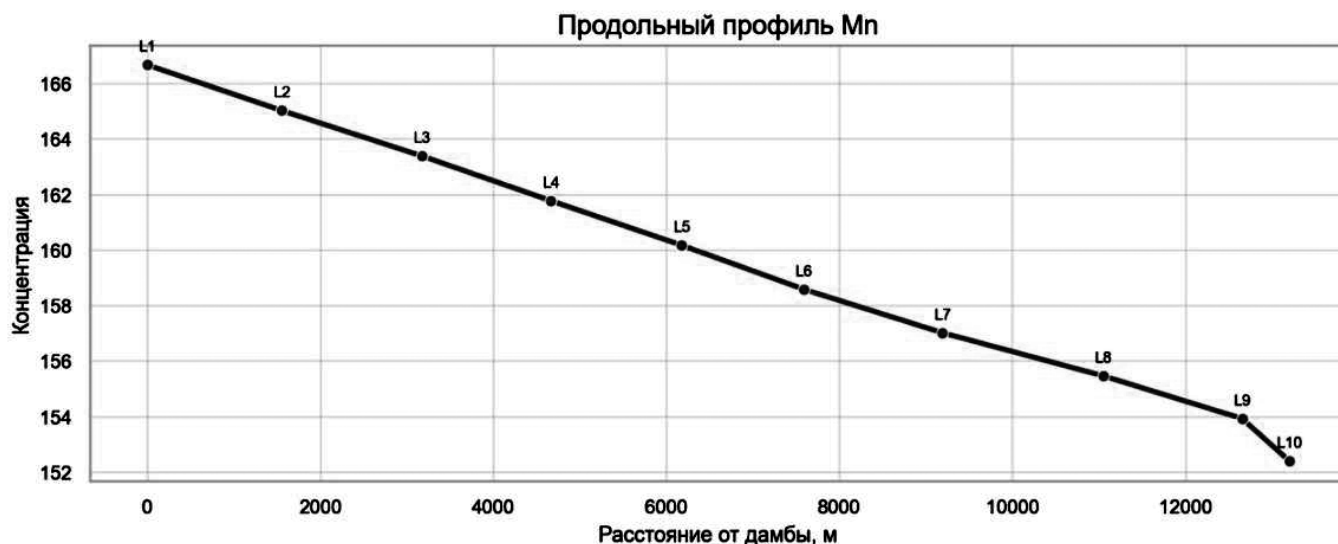


Рис. 3. Распределение с применением логарифмической шкалы (первый вариант визуализации)

на рис. 2–3. Его преимущество 0 простота и наглядность при создании единой карты интегрального индекса загрязнения.

Второй вариант (многослойное наложение) представлен на рис. 4–6. При совмещении слоёв (рис. 4) зоны совместного присутствия обоих металлов отображаются фиолетовыми оттенками, зоны преобладания марганца — красными, меди — синими. Такое представление позволяет визуально разделить вклад каждого загрязнителя.

На рис. 5 показано пространственное поле превышений ПДК для марганца. Максимальные значения (более 50 ПДК) приурочены к участку непосредственно ниже места прорыва дамбы. Поле превышений ПДК для меди

(рис. 6) имеет схожую структуру, однако абсолютные значения ниже (не более 10 ПДК), что обусловлено меньшим содержанием меди в отходах обогащения.

Второй вариант более трудоёмок, но он незаменим при детальном расследовании, когда необходимо оценить вклад каждого поллютанта.

4. Анализ динамики загрязнения по временным периодам

Для изучения изменения экологической ситуации после аварии построены карты интегрального индекса загрязнения для трёх временных интервалов (рис. 7–9).

В фоновый период (рис. 7) превышения ПДК носят локальный характер и приурочены в основном к зоне



Рис. 4. Интегральная карта загрязнения, полученная наложением слоёв марганца (красный канал) и меди (синий канал)

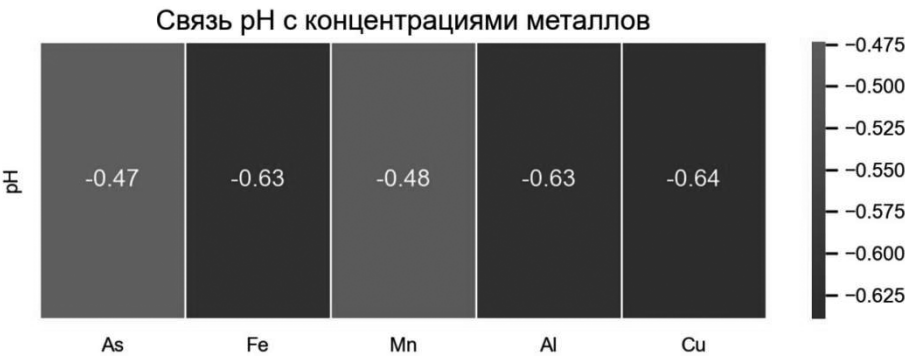


Рис. 5. Распределение концентрации марганца (в единицах ПДК)



Рис. 6. Распределение концентрации меди (в единицах ПДК)

хвостохранилища, что связано с плановыми сбросами оборотных вод. Индекс загрязнения не превышает 5–10 единиц.

В течение года после аварии (рис. 8) фиксируется многократное (до 50–100 раз) превышение ПДК на всём

протяжении водотока ниже дамбы. Зона загрязнения распространяется на десятки километров, что свидетельствует о масштабном разовом выбросе.

В отдалённый период (рис. 9) наблюдается постепенное снижение концентраций по мере удаления

```
# базовая разметка
step_cmp_01_data["Period"] = "Before accident"

# острый период (1 год после аварии)
step_cmp_01_data.loc[
    (step_cmp_01_data["Date"] >= "2021-07-03") &
    (step_cmp_01_data["Date"] <= "2022-06-30"),
    "Period"
] = "Acute phase"

# пост-аварийный период
step_cmp_01_data.loc[
    step_cmp_01_data["Date"] > "2022-06-30",
    "Period"
] = "Post-accident"
```

Рис. 7. Пространственное распределение интегрального индекса загрязнения в фоновый период (2002 — 2 июля 2021 г.)

от источника. Однако вблизи хвостохранилища индекс загрязнения остаётся повышенным, что указывает на продолжающееся вторичное загрязнение за счёт размыва донных отложений.

5. Анализ превышений ПДК вдоль продольного профиля

Для количественной оценки пространственного распределения загрязнения построены графики изменения отношения концентрации к ПДК вдоль водотока (рис. 10–12). На рис. 10 видно, что в острый период пиковые значения (до 100 ПДК для марганца) приходятся на зону до 10 км от прорыва, затем наблюдается постепенное снижение. В фоновый период превышения не превышают 5 ПДК. В отдалённый период значения снижаются, но остаются выше фоновых на всём протяжении.

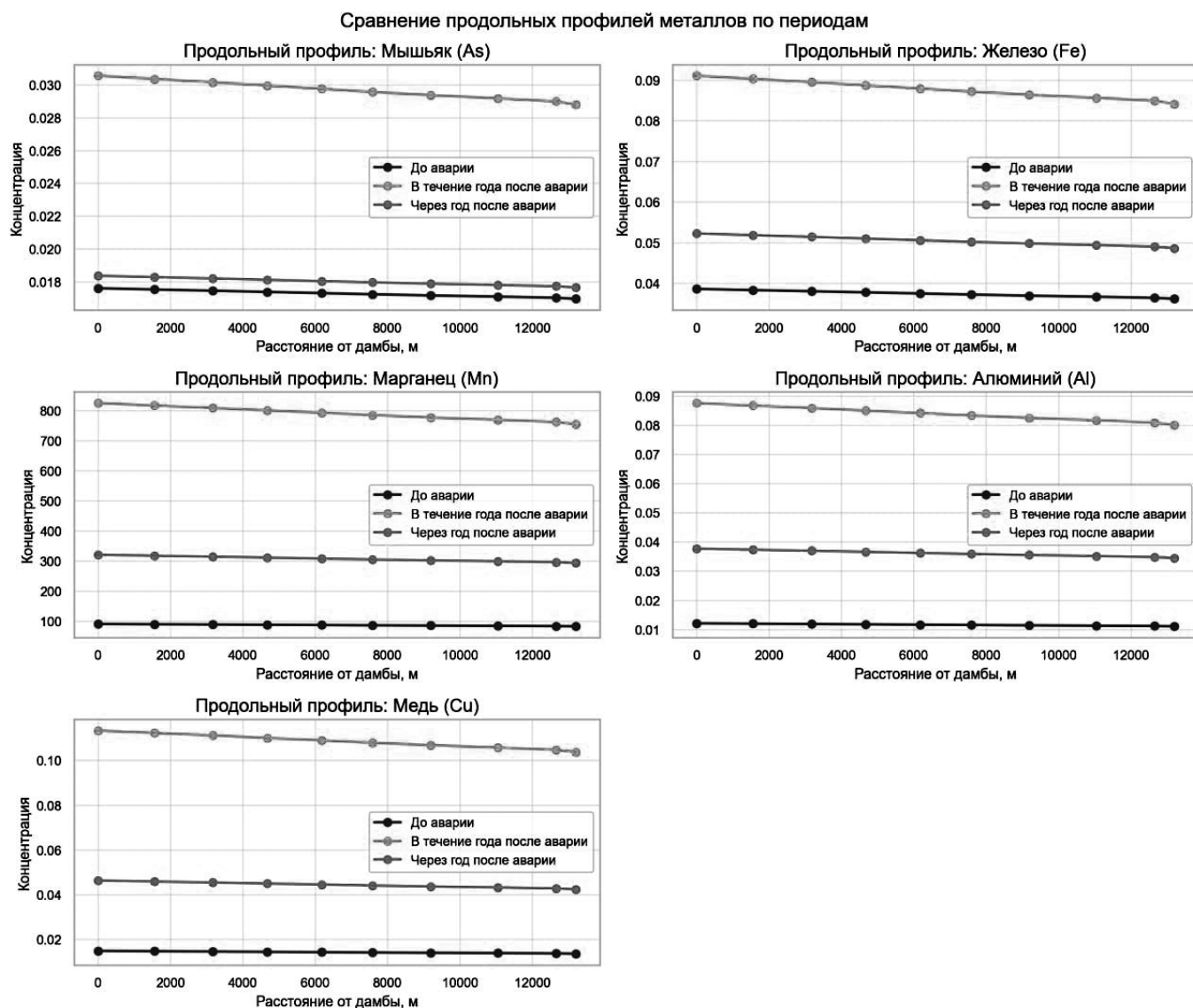


Рис. 8. График показателей острого послеаварийного периода (3 июля 2021 г. — 30 июня 2022 г.)

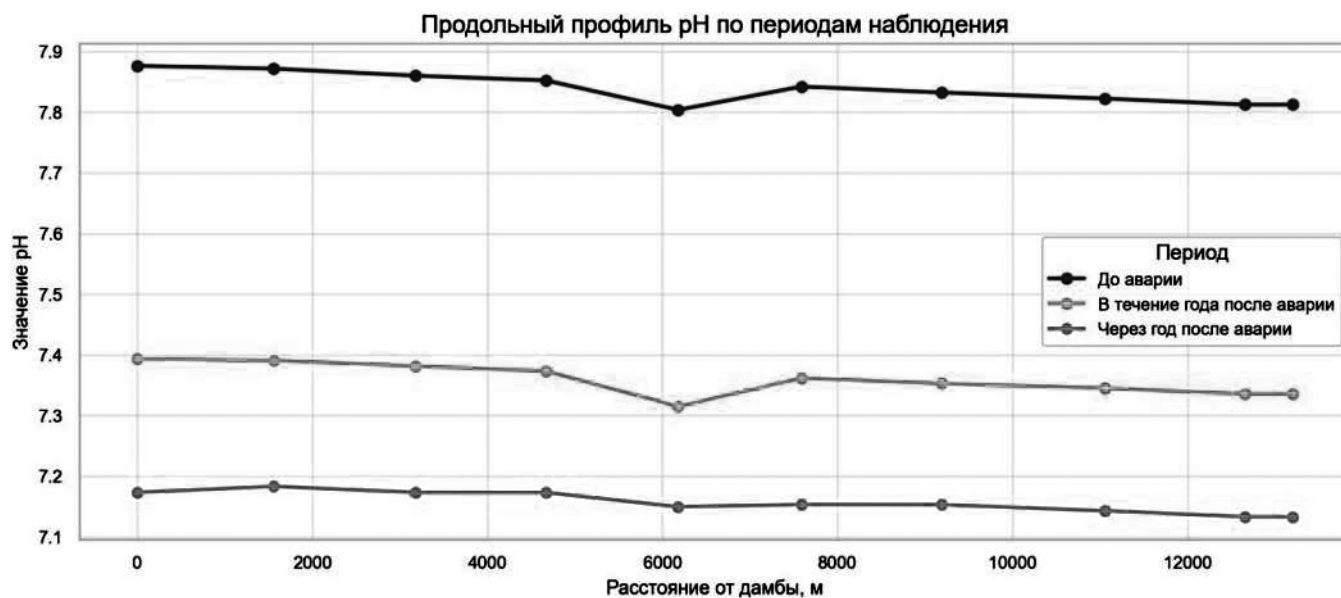


Рис. 9. График показателей отдалённого периода (после 30 июня 2022 г.)

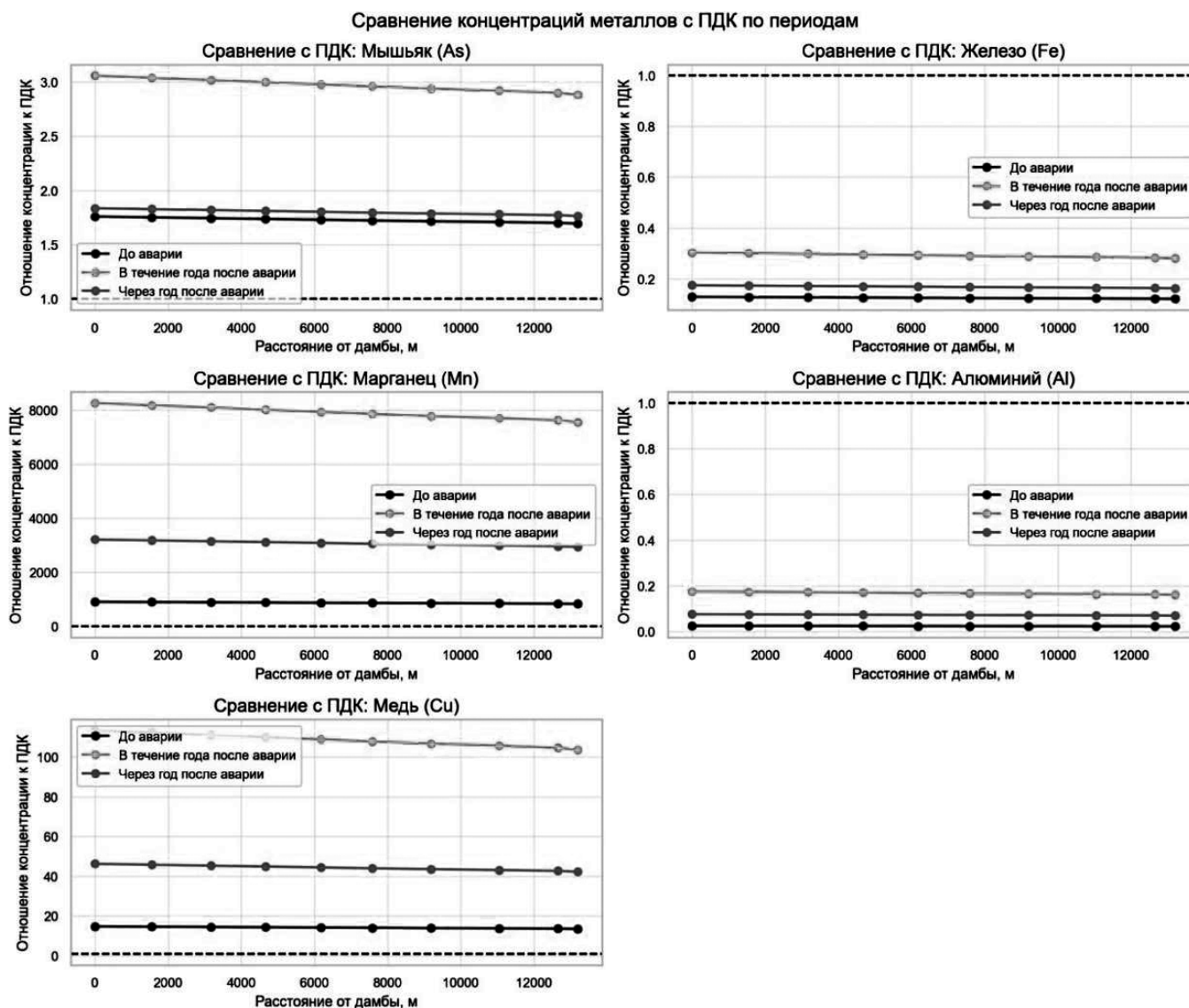


Рис. 10. Изменение превышений ПДК для марганца и меди по мере удаления от дамбы (для трёх периодов)

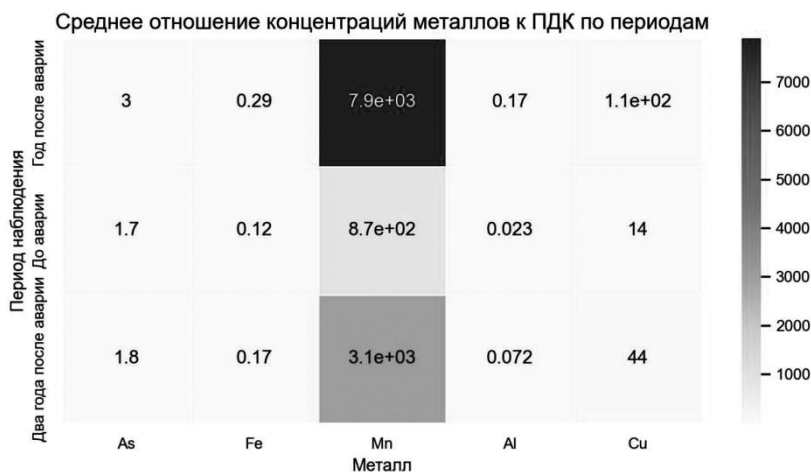


Рис. 11. Динамика отношения концентрации марганца к ПДК

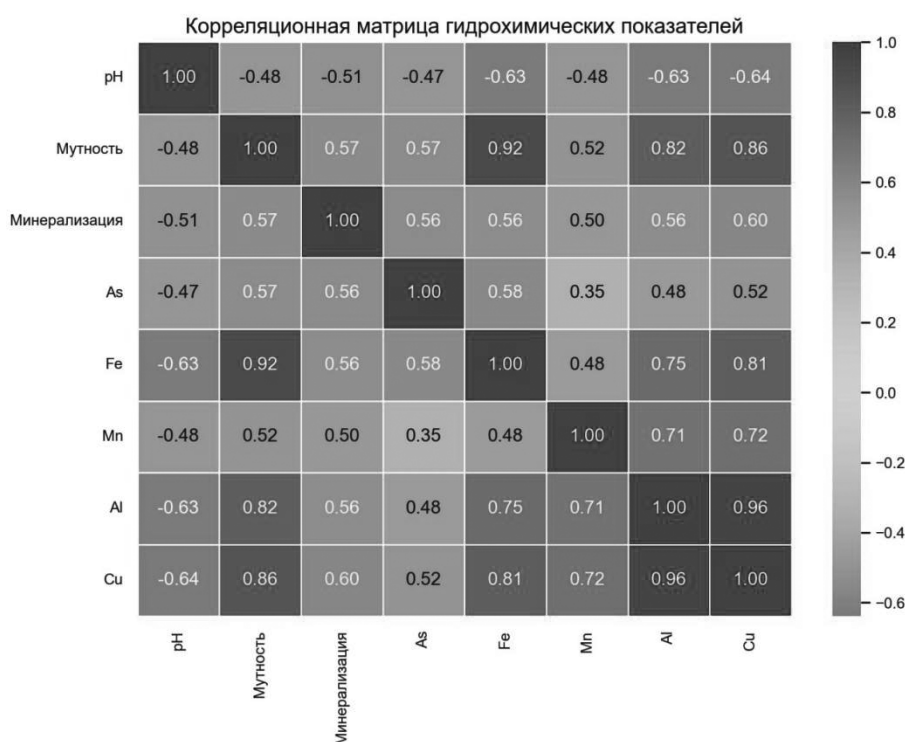


Рис. 12. Динамика отношения концентрации меди к ПДК

Из-за крайне низкой величины ПДК для марганца (0,01 мг/л) даже относительно небольшие концентрации дают высокие индексы загрязнения (рис. 11). Это объясняет, почему марганец является основным индикатором техногенного воздействия в рассматриваемом районе.

Медь характеризуется менее выраженными пиками (максимум до 10 ПДК), но сохраняет превышения на значительном удалении (до 30 км), что свидетельствует о её более высокой миграционной способности (рис. 12).

Полученные результаты подтверждают эффективность предложенных методов визуализации и автоматизированной обработки спутниковых данных для экологического мониторинга и расследования последствий

аварий на хвостохранилищах [8]. Выявленные закономерности могут быть использованы при планировании природоохранных мероприятий и в системах поддержки принятия решений.

Заключение

В ходе исследования решены следующие задачи:

1. Разработана и реализована автоматизированная технология обработки многолетних (более 20 лет) спутниковых данных для района хвостохранилища Катока (Ангола) с использованием облачных вычислений.
2. Предложены и апробированы два подхода к визуализации интегрального индекса загрязнения:

логарифмическое масштабирование и многослойное наложение, позволяющие корректно отображать широкий диапазон концентраций.

3. Выполнен анализ динамики загрязнения марганцем и медью до и после аварии; установлено, что в острый период превышения ПДК достигали 100-кратных значений, а в отдалённый период наблюдается частичное самоочищение, однако сохраняется хроническое загрязнение.
4. Показано, что сочетание спутникового мониторинга и автоматизированных методов обработки

данных создаёт объективную основу для экологического расследования техногенных аварий и может быть масштабировано на другие горнорудные регионы.

Разработанные методики могут быть использованы при создании автоматизированных систем экологического контроля и поддержки принятия решений в чрезвычайных ситуациях.

ЛИТЕРАТУРА

1. Rico M., Benito G., Díez-Herrero A. Floods from tailings dam failures // *Journal of Hazardous Materials*. — 2008. — Vol. 154, Iss. 1–3. — P. 79–87.
2. Кронберг П. Дистанционное изучение Земли: основы и методы дистанционных исследований в геологии. — М.: Мир, 1988. — 343 с.
3. Берлянт А.М. Картография: учебник. — М.: КДУ, 2011. — 464 с.
4. Chuvieco E. *Fundamentals of Satellite Remote Sensing: An Environmental Approach*. — CRC Press, 2016. — 468 p.
5. ГН 2.1.5.1315-03. Предельно допустимые концентрации (ПДК) химических веществ в воде водных объектов хозяйственно-питьевого и культурно-бытового водопользования. — М.: Минздрав России, 2003.
6. Tufte E.R. *The Visual Display of Quantitative Information*. — Graphics Press, 2001. — 197 p.
7. Кравцова В.И. Космические методы картографирования. — М.: Изд-во МГУ, 1995. — 240 с.
8. Santos R., Andrade C., Guerra M. et al. Remote sensing for environmental monitoring of mining areas: A case study in Angola // *International Journal of Applied Earth Observation and Geoinformation*. — 2022. — Vol. 112. — 102934.

© Тшипунгу Эдуарду Афонсу (tchipungu369@mail.ru); Шатунова Вероника Максимовна

Журнал «Современная наука: актуальные проблемы теории и практики»

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ ЭНЕРГОЭФФЕКТИВНЫХ РЕЖИМОВ КРИОГЕННОГО РАЗДЕЛЕНИЯ ВОЗДУХА В УСЛОВИЯХ ТЕХНОЛОГИЧЕСКОЙ НЕОПРЕДЕЛЕННОСТИ

MATHEMATICAL MODELING AND OPTIMIZATION OF ENERGY- EFFICIENT CRYOGENIC AIR SEPARATION MODES UNDER CONDITIONS OF TECHNOLOGICAL UNCERTAINTY

R. Uzbekov
V. Chelnokov

Summary. This paper presents an integrated approach to the construction of a mathematical model of a distillation column as a control object. A systematic analysis of the sources of technological uncertainty characteristic of large-scale industries has been carried out. Modern methods of nonlinear optimization are considered, including deterministic (sequential quadratic programming methods), stochastic (genetic algorithms, swarm optimization) and hybrid schemes. The architecture of an optimization system based on a digital twin is proposed, integrating a machine learning unit to identify uncertainty parameters.

Keywords: cryogenic air separation, rectification, mathematical modeling, energy efficiency, nonlinear optimization, technological uncertainty, digital twin, hybrid algorithms.

Узбеков Руслан Шавкятъевич

Аспирант, Российский химико-технологический
университет им. Д.И. Менделеева, г. Москва
uzbekovruslan01@yandex.ru

Челноков Виталий Вячеславович

доктор технических наук, профессор,
Российский химико-технологический
университет им. Д.И. Менделеева, г. Москва

Аннотация. В настоящей работе представлен комплексный подход к построению математической модели колонны ректификации как объекта управления. Проведен системный анализ источников технологической неопределенности, характерных для крупнотоннажных производств. Рассмотрены современные методы нелинейной оптимизации, включая детерминированные (методы последовательного квадратичного программирования), стохастические (генетические алгоритмы, роевая оптимизация) и гибридные схемы. Предложена архитектура системы оптимизации на основе цифрового двойника, интегрирующая блок машинного обучения для идентификации параметров неопределенности.

Ключевые слова: криогенное разделение воздуха, ректификация, математическое моделирование, энергоэффективность, нелинейная оптимизация, технологическая неопределенность, цифровой двойник, гибридные алгоритмы.

Введение

Криогенные процессы разделения воздуха методом низкотемпературной ректификации остаются доминирующим технологическим решением для получения больших объемов высокочистых кислорода, азота и аргона. Некриогенные методы (адсорбция при переменном давлении, мембранное разделение), несмотря на развитие, не могут конкурировать с криогенной дистилляцией в крупнотоннажном секторе [1]. Однако основным недостатком таких производств является исключительно высокая энергоемкость: годовое электропотребление одной установки производительностью 50 000 Нм³/ч кислорода может превышать 400 ГВт/ч.

В условиях роста тарифов на энергоносители и ужесточения экологических норм задача повышения энергоэффективности становится критической для сохранения конкурентоспособности продукции. Традиционные методы поиска оптимальных режимов, основанные

на эмпирических правилах или локальной линеаризации, часто оказываются неспособны учесть комплексные взаимовлияния параметров в нелинейной системе «колонна — теплообменное оборудование — компрессорная группа» [2–3]. Ситуация усугубляется наличием технологической неопределенности, которая делает детерминированные оптимизационные решения неоптимальными или даже нереализуемыми.

Целью данной работы является разработка методологии математического моделирования и оптимизации энергоэффективных режимов криогенного разделения воздуха, учитывающей нелинейный характер процессов и множество факторов неопределенности.

Исследование

Технологическая схема разделения воздуха.

Процесс разделения воздуха основан на различии температур кипения его компонентов: азот (N₂) кипит

при $-195,8^{\circ}\text{C}$, аргон (Ar) при $-185,9^{\circ}\text{C}$, кислород (O_2) при $-183,0^{\circ}\text{C}$ [4]. Для достижения высокой чистоты продуктов промышленные установки используют колонну двукратной ректификации, включающую нижнюю колонну высокого давления (КВД) и верхнюю колонну низкого давления (КНД) [5–6]. Технологическая схема включает следующие основные этапы:

1. Подготовка воздуха: сжатие в главном воздушном компрессоре (ГВК) до 5–6 бар, очистка от влаги, CO_2 и углеводородов в адсорберах, охлаждение в основном теплообменнике-рекуператоре.
2. Предварительное разделение в КВД: воздух подается в нижнюю колонну, где в результате ректификации получают газообразный азот высокой чистоты (вверх) и кубовую жидкость, обогащенную кислородом (до 30–40 % O_2).
3. Окончательное разделение в КНД: после дросселирования кубовая жидкость поступает в верхнюю колонну (давление 1,3–1,5 бар), где происходит отделение чистого кислорода (в кубе) и азота (вверх). При необходимости получения аргона производится отбор аргонной фракции из средней части КНД и ее очистка в отдельной аргонной колонне.

Важно отметить, что термодинамическая эффективность всей установки определяется согласованной работой обеих ступеней ректификации и системой рекуперации холода. Любое отклонение режима работы одной колонны влияет на тепловой баланс другой, что создает сложную систему перекрестных связей [7].

Анализ современных исследований и производственных данных позволяет систематизировать источники неопределенности, влияющие на эффективность воздуходелительной установки (ВРУ) [8]:

1. Неопределенность сырья (атмосферного воздуха):

Колебания состава (влажность, содержание CO_2 , наличие углеводородов и закиси азота N_2O); изменение термодинамических параметров (давление, температура) в зависимости от климатических условий и времени года; загрязненность промышленными выбросами, что приводит к изменению эффективности адсорберов и теплообменников.

2. Эксплуатационная неопределенность:

Изменчивость нагрузки — колебания потребности в продуктах (O_2 , N_2 , Ar) со стороны производства; динамика стоимости электроэнергии, что требует переключения режимов для минимизации затрат; режимы пуска, останова и переходные процессы, которые могут занимать до 10 % времени эксплуатации.

3. Аппаратурно-параметрическая неопределенность:

Неточность моделей тепло- и массообмена, используемых при проектировании; неопределенность теплофизических свойств материалов (теплопроводность изоляции, эффективность насадок); дрейф характеристик оборудования во времени: загрязнение теплообменников (снижение коэффициента теплопередачи на 1–2 % в год), износ детандеров, изменение гидравлического сопротивления тарелок.

4. Внешняя неопределенность:

Рыночные колебания (спрос/цена на продукты); изменения нормативной базы (экологические и промышленные стандарты); климатические воздействия (суточные и сезонные колебания температуры), влияющие на работу компрессоров и конденсаторов.

Данные факторы приводят к тому, что оптимальный режим, рассчитанный для номинальных условий, перестает быть таковым в реальной эксплуатации. Игнорирование неопределенности может привести к нарушению технологических ограничений (например, падение чистоты продукта) или к перерасходу энергии.

Структура математической модели ВРУ

Создание адекватной математической модели — ключевой этап разработки системы оптимизации. Согласно принципам системного анализа, воздуходелительная установка представляет собой иерархическую структуру с многоуровневыми связями [9]. Нижний уровень ($L=0$) образуют отдельные аппараты (ректификационные колонны, теплообменники, детандеры), уровень $L=1$ — подсистемы разделения, охлаждения, сжатия, и уровень $L=2$ — установка в целом.

Для задач оптимизации энергоэффективности целесообразно использовать модель «вход-выход» с закрытой внутренней структурой:

- Входные управляющие переменные (U): расход воздуха на входе, распределение нагрузок между компрессорами, давление в колоннах, расходы флегмы и орошения;
- Выходные переменные (Y): чистота продуктов, расход продуктов, удельное энергопотребление ($\text{кВт}\cdot\text{ч}/\text{Нм}^3$);
- Возмущающие воздействия (F): температура и влажность воздуха на входе, температура охлаждающей воды, состав сырья.

Нелинейность моделей криогенных установок обусловлена фундаментальными физико-химическими законами. Термодинамические свойства компонентов

воздуха в рабочем диапазоне температур (от -196°C до $+30^{\circ}\text{C}$) описываются сложными уравнениями состояния (например, модифицированное уравнение Бенедикта-Уэбба-Рубина) [10], которые являются существенно нелинейными и неявными относительно энтальпии и энтропии.

Для моделирования ректификационных колонн используются ступенчатые модели, основанные на уравнениях материального и теплового баланса для каждой теоретической тарелки. Для i -й тарелки и j -го компонента уравнения имеют вид:

- Материальный баланс:

$$L_{i-1}x_{i-1,j} + V_{i+1}y_{i+1,j} - L_i x_{i,j} - V_i y_{i,j} = 0$$

- Уравнение фазового равновесия:

$$y_{i,j} = K_{i,j}(T_i, P_i, x_{i,j}) \cdot x_{i,j}$$

- Сумма мольных долей:

$$\sum_j x_{i,j} = 1, \sum_j y_{i,j} = 1$$

Здесь L и V — потоки жидкости и пара, x и y — мольные доли в жидкой и паровой фазах, K — коэффициент фазового равновесия, который сам является нелинейной функцией температуры, давления и состава.

Гидродинамические процессы в колоннах (потери на тарелках, захлебывание) описываются зависимостями, включающими нелинейные функции от плотности и скорости потоков [11]. Характеристики центробежных компрессоров и детандеров представляют собой семейства нелинейных кривых, аппроксимируемых полиномами высоких степеней. Все это приводит к тому, что целевая функция (удельное энергопотребление) становится мультимодальностью.

Построение модели осложняется недоступностью измерения многих внутренних переменных. В этом случае применяется принцип последовательного раскрытия неопределенности [12]. Процедура идентификации включает:

1. Определение структуры операторов. Как правило, принимается гипотеза о том, что каналы могут быть описаны устойчивыми звеньями первого порядка с запаздыванием:

$$G(s) = \frac{K}{Ts + 1} e^{-\tau s}$$

где K — коэффициент усиления, T — постоянная времени (определяет инерционность системы), τ — время запаздывания, s — оператор Лапласа.

2. Пассивную идентификацию — анализ архивных данных за длительный период (фильтрация шумов, исключение экстремальных значений) с использованием методов регрессионного анализа и нейронных сетей.

3. Активную идентификацию — проведение специальных экспериментов на действующей установке в допустимых диапазонах изменения переменных.

Критически важным этапом является калибровка модели по реальным эксплуатационным данным для учета старения оборудования и погрешностей измерительных каналов. Для этого применяются методы оптимизации (например, минимизация среднеквадратичного отклонения между расчетными и измеренными значениями ключевых параметров: чистоты продуктов, перепадов давления, температур). Адекватность модели проверяется путем сравнения расчетных и фактических данных на независимой выборке.

Методы оптимизации энергоэффективных режимов

Задача оптимизации формулируется как поиск минимума целевой функции удельного энергопотребления $F(X)$ при соблюдении технологических ограничений (чистота продуктов не ниже заданной, не превышение допустимых нагрузок оборудования). Вектор управляющих переменных X включает давления в колоннах, расходы отборов, распределение нагрузки между параллельно работающими компрессорами, расходы флегмы.

Для задач с нелинейными ограничениями наиболее эффективны методы последовательного квадратичного программирования (SQP) [13]. На каждой итерации исходная задача аппроксимируется задачей квадратичного программирования: строится квадратичная аппроксимация функции Лагранжа и линеаризуются ограничения.

Алгоритм SQP:

1. Выбрать начальную точку x_0 ;
2. Вычислить градиенты $\nabla F(x_k)$ и $\nabla g_j(x_k)$, а также матрицу Гессе функции Лагранжа (или ее аппроксимацию);
3. Решить подзадачу квадратичного программирования для определения направления поиска d_k ;
4. Определить длину шага α_k (например, методом золотого сечения);
5. Обновить точку: $x_{k+1} = x_k + \alpha_k d_k$
6. Проверить критерий сходимости; если не выполнен, перейти к шагу 2.

Преимущества SQP — высокая скорость сходимости в окрестности экстремума и точность. Недостатки — чувствительность к начальному приближению и возможность схождения к локальному, а не глобальному минимуму.

Для глобального поиска в многомодальном пространстве целесообразно применять генетические алгоритмы (ГА) и оптимизацию роем частиц (PSO).

Генетические алгоритмы работают с популяцией решений, закодированных в виде «хромосом» (наборов параметров). Основные этапы [14]:

- Инициализация случайной популяции;
- Оценка приспособленности (вычисление целевой функции);
- Селекция лучших особей;
- Кроссинговер (обмен участками хромосом) и мутация (случайные изменения);
- Формирование нового поколения.

Преимущество ГА — робастность и способность «перепрыгивать» локальные минимумы благодаря механизмам мутации.

Метод роевой оптимизации (PSO) имитирует поведение стаи: каждая частица (кандидатное решение) движется в пространстве поиска, корректируя траекторию на основе собственного лучшего положения (*pbest*) и лучшего положения роя в целом (*gbest*) [15]. Скорость частицы обновляется по формуле:

$$v_i^{t+1} = wv_i^t + c_1r_1(pbest_i - x_i^t) + c_2r_2(gbest - x_i^t)$$

где w — инерционный вес, c_1, c_2 — коэффициенты ускорения, r_1, r_2 — случайные числа. PSO проще в настройке и эффективен для гладких многомерных задач.

Наиболее перспективным направлением признается двухэтапная гибридная схема:

1. Этап глобального поиска. Используется метаэвристика (ГА или PSO) для грубого исследования всего пространства параметров и идентификации области, содержащей глобальный оптимум. Обычно достаточно 50–100 поколений (для ГА) или 200–300 итераций (для PSO).
2. Этап локальной оптимизации. Найденное на первом этапе решение используется как начальное приближение для быстрого градиентного метода (SQP), который точно определяет координаты оптимума с заданной точностью (например, до 0,01 % по целевой функции) [16].

Такая комбинация позволяет преодолеть недостатки каждого из методов в отдельности и гарантированно находить глобальный экстремум для ВПУ за приемлемое время.

Многокритериальная оптимизация

В ряде случаев требуется одновременно минимизировать энергопотребление и максимизировать выход

аргона или надежность работы. Такие задачи решаются методами многокритериальной оптимизации, например, с использованием взвешенной суммы критериев или поиска Парето-фронта. Генетические алгоритмы (NSGA-II) хорошо подходят для построения Парето-множества [17].

Учет неопределенности в оптимизационной модели может быть реализован через:

- Робастную оптимизацию — поиск решения, гарантирующего выполнение ограничений при любых реализациях неопределенных параметров в заданном множестве. Однако робастные решения могут быть излишне консервативными (потери энергоэффективности до 5 %).
- Стохастическое программирование — оптимизация математического ожидания целевой функции с учетом вероятностных распределений неопределенных параметров. Требуется знания законов распределения, что на практике не всегда доступно.
- Адаптивные схемы с обучением — использование методов машинного обучения (кластеризация, байесовские методы, оценка плотности распределения) для извлечения актуальной информации о неопределенности из текущих данных и адаптации модели в реальном времени. Например, на основе данных о температуре воздуха за последние сутки можно пересчитать оптимальные уставки давления в колоннах.

Основой внедрения оптимизационных алгоритмов является создание цифрового двойника технологической установки [18]. Цифровой двойник включает:

- Нелинейную статическую и динамическую модель установки (реализованную в среде Aspen Plus, HYSYS или MATLAB/Simulink);
- Модуль идентификации и калибровки на основе данных АСУ ТП (с использованием нейронных сетей или регрессионного анализа);
- Блок оптимизации, реализующий выбранные алгоритмы (SQP, ГА, PSO или гибридные);
- Интерфейс передачи уставок в систему управления (OPC-сервер, Modbus).

Возможны два основных режима работы, онлайн-оптимизация — система периодически (например, каждый час) решает задачу оптимизации и автоматически корректирует уставки регуляторов (расходы, давления). Требуется высокой надежности модели и систем сбора данных. Также это офлайн-режим, который используется технологическим персоналом для анализа сценариев, планирования режимов, обучения и поддержки принятия решений (например, что будет, если изменить нагрузку на 10 %).

Внедрение системы оптимизации требует:

- Интеграции с существующей АСУ ТП (SCADA-система, программируемые логические контроллеры);
- Обеспечения вычислительных мощностей (сервер для расчетов);
- Разработки защит и блокировок, исключающих передачу аварийных уставок;
- Обучения персонала.

Особое внимание уделяется фильтрации входных данных и отбраковке выбросов (например, при сбоях датчиков), чтобы неверные данные не приводили к некорректным рекомендациям.

Согласно опубликованным данным по внедрению систем оптимизации на криогенных ВРУ, достигаемое снижение удельного энергопотребления составляет от 2 % до 7 % без капитальных затрат на модернизацию оборудования [19]. Для установки производительностью 50 000 Нм³/ч кислорода годовая экономия электроэнергии может превысить 25 ГВт/ч, что эквивалентно

экономии более 100 млн рублей в год (при тарифе 4 руб. за кВт/ч). Срок окупаемости системы оптимизации обычно не превышает 1–1,5 года. Дополнительный эффект дает повышение стабильности качества продукции (чистоты) и снижение числа нарушений технологического режима.

Заключение

Проведенный анализ подтверждает, что повышение энергоэффективности криогенных производств требует нелинейных моделей, учитывающих неопределенность процессов. Наиболее перспективны гибридные алгоритмы оптимизации. Ключевое условие — цифровой двойник, интеграция с АСУ ТП обеспечивает снижение энергопотребления. Дальнейшее развитие связано с интеграцией методов нелинейной оптимизации и машинного обучения для создания адаптивных систем реального времени. Также актуально развитие методов многокритериальной оптимизации для поиска оптимального решения между энергоэффективностью, надежностью, экологичностью и гибкостью производства.

ЛИТЕРАТУРА

1. Результаты исследований по идентификации причин нестабильной работы криогенной воздухоразделительной установки ТКДС-100В / А.А. Воробьев, А.В. Козлов, А.М. Кокарев, И.В. Желтоухов // Вестник Московского государственного технического университета им. Н. Э. Баумана. Серия «Машиностроение». — 2022. — Т. 4, № 143. — С. 89–107.
2. Райхлин, С.М. Экономическая оценка проектов повышения энергоэффективности на горных предприятиях в целях устойчивого развития: специальность 5.2.3. «Региональная и отраслевая экономика»: диссертация на соискание ученой степени кандидата экономических наук / Райхлин Семен Максимович. — Санкт-Петербург, 2024. — 134 с.
3. Редникова, Т.В. Правовое регулирование энергоэффективности и энергосбережения в России и за рубежом: экологический вектор / Т.В. Редникова // Ученые записки Казанского университета. Серия Гуманитарные науки. — 2025. — Т. 167, № 5–6. — С. 282–297.
4. Мостовенко, Л.В. Основы промышленной теплоэнергетики: учебное пособие / Л.В. Мостовенко, В.П. Белоглазов. — Нижневартовск: Нижневартовский государственный университет, 2022. — 124 с.
5. Соколова, Е.В. Основы проектирования низкотемпературной техники: учебно-методическое пособие / Е.В. Соколова. — Санкт-Петербург: Университет ИТМО, 2022. — 43 с.
6. Луканин, П.В. Низкотемпературные процессы и установки: учебное пособие / П.В. Луканин, Г.А. Морозов. — Санкт-Петербург: ВШТЭ СПбГУПТД, 2022. — 135 с.
7. Анализ экономической эффективности использования рабочего тела в кон туре регазификации, работающем на основе закрытого цикла Ренкина / Е.В. Благин, Д.А. Угланов, В.В. Урлапкин [и др.] // Холодильная техника. — 2024. — Т. 113, № 4. — С. 188–199.
8. Кондратьев, В.Б. Роль критически важных сырьевых материалов в условиях экономической неопределенности: опыт ЕС / В.Б. Кондратьев // Горная промышленность. — 2022. — № 4. — С. 94–102.
9. Плотникова Л.В., Торкунова Ю.В. Программное обеспечение системного анализа сложноструктурированных промышленных комплексов при разработке энергосберегающих мероприятий // Вестник Казанского государственного энергетического университета. — 2022. — Т. 14. — №. 1 (53). — С. 140–154.
10. Колдоба Е.В. Численное решение многопараметрического уравнения состояния Бенедикта-Вебба-Рубина // Интеллектуальные системы. Теория и приложения. — 2024. — Т. 28. — №. 2. — С. 5–11.
11. Лаптева Е.А., Дударовская О.Г. Математическая модель процесса абсорбции в насадочных колоннах // Инженерный вестник Дона. — 2022. — №. 4 (88). — С. 105–115.
12. Кораблёв Ю.А., Лосева Д.М., Шестопапов М.Ю. Концепция эволюционного синтеза современных сложных систем управления // Известия СПбГЭТУ ЛЭТИ. — 2022. — Т. 34. — №4. С. 29.
13. Никулин А.С., Куркин Д.С., Жедяевский Д.Н. Оптимизация энергетических затрат при производстве сжиженного природного газа в арктических условиях // Автоматизация и информатизация ТЭК. — 2022. — №. 7 (588). — С. 5.
14. Душина В.А. Генетические алгоритмы для задачи оптимизации мультимодальной функции // StudNet. — 2022. — Т. 5. — №. 7. — С. 7824–7848.
15. Казакова Е.М. Краткий обзор методов оптимизации на основе роя частиц // Вестник КРАУНЦ. Физико-математические науки. — 2022. — Т. 39. — №. 2. — С. 150–174.

16. Никулин А.С., Жедяевский Д.Н., Федорова Е.Б. Применение искусственных нейронных сетей для подбора состава смесового хладагента с заданной кривой кипения //Компьютерные исследования и моделирование Учредители: Институт компьютерных исследований, Университет Иннополис. — 2022. — Т. 14. — №. 3. — С. 593–608.
17. Шырын Б.А., Жумадилаева А.К., Бекешова Г.Б. Многоцелевая эволюционная оптимизация политик безопасности в программно-определяемых сетях (sdn) с учетом ограничений tsat и задержки //Вестник Университета Шакарима. Серия технические науки. — 2025. — №. 4. — С. 238–248.
18. Лозница С.Ю., Лапшин Ю.А., Кроль С.П. Математическая модель радиуса разворота самолёта с учётом ветра на основе алгоритма муравьиной колонии и концепции цифрового двойника //Парадигма. — 2025. — №. 11-4. — С. 101–106.
19. Кондратенко А.Н. Вопросы методологии сопровождения, реализации и информационного обеспечения научно-исследовательских и опытно-конструкторских работ, технологических работ и инвестиционных проектов //Вестник НПО Техномаш. — 2022. — №. 1 (18). — С. 28.

© Узбеков Руслан Шавкятъевич (uzbekovruslan01@yandex.ru); Челноков Виталий Вячеславович
Журнал «Современная наука: актуальные проблемы теории и практики»

РАЗРАБОТКА СИСТЕМЫ СИМУЛЯЦИИ РАБОТЫ АЛГОРИТМОВ УПРАВЛЕНИЯ ЭЛЕКТРОЭНЕРГИЕЙ В МУЛЬТИАГЕНТНЫХ СИСТЕМАХ В РАМКАХ МИКРОГРИДА НА БАЗЕ МОДЕЛИ ДВОЙНОГО АУКЦИОНА

Цай Цин

Аспирант, Университет ИТМО
caiq041@gmail.com

DEVELOPMENT OF A SIMULATION SYSTEM FOR ENERGY MANAGEMENT ALGORITHMS IN MULTI-AGENT SYSTEMS WITHIN A MICROGRID BASED ON THE DOUBLE AUCTION MODEL

Tsai Tsin

Summary. The objective is to develop a software environment for the simulation of energy management algorithms based on the double auction market mechanism. The model distinguishes four types of economic agents: consumers, generators powered by renewable energy sources, energy storage systems, and a market agent performing the role of a centralized auctioneer. For each agent type, an objective function and bid formation logic are defined. The computational experiments confirm the operability of the algorithm: a statistically significant negative correlation between price and system balance is observed, and the efficiency of the price arbitrage strategy executed by storage agents is demonstrated. The resulting software environment is intended for further research, including the integration of large language models into agent decision-making logic.

Keywords: energy management algorithm, distributed energy network, microgrid, double auction, multi-agent system, energy consumption optimization, SPADE.

Аннотация. Целью работы является разработка программной среды для имитационного моделирования алгоритмов управления электроэнергией, основанных на рыночном механизме двойного аукциона. В модели выделены экономические агенты четырёх типов: потребители, генераторы на основе возобновляемых источников, системы накопления и агент рынка, выполняющий функцию централизованного аукциониста. Для каждого типа агентов сформулирована целевая функция и описана логика формирования рыночной заявки. По результатам зафиксирована статистически значимая отрицательная корреляция между ценой и балансом энергосистемы, а также эффективность стратегии ценового арбитража накопителей. Полученная программная среда предназначена для дальнейших исследований.

Ключевые слова: алгоритм управления электроэнергией, распределённая энергосеть, микрогрид, двойной аукцион, мультиагентная система, оптимизация энергопотребления, SPADE.

Введение

Микросети, или микрогриды (англ. microgrids), объединяют возобновляемые источники энергии, нагрузки и системы хранения. Управление в подобных системах осложнено непостоянством возобновляемой генерации и отсутствием достаточных резервов для балансировки. Традиционные системы SCADA требуют значительных затрат на внедрение в условиях микросетей и ограничены в совместимости устройств. Перспективную альтернативу открывает децентрализованное управление отдельными компонентами с помощью автономных программных агентов, балансирующих генерацию и спрос [4–7].

Наиболее распространённым инструментом построения мультиагентных систем для управления энергопо-

треблением и третичного контроля в микросетях до недавнего времени оставался фреймворк JADE [8]. Однако он обладает рядом ограничений, связанных с особенностями используемой платформы Java и протокола Java Messaging Service. Фреймворк SPADE (Smart Python Agent Development Environment), основанный на Python [1–3], поддерживает протоколы XMPP, HTTP и TCP для организации коммуникации между агентами при их децентрализованном взаимодействии. Указанные свойства делают его пригодным для систем Интернета вещей и встроенных приложений и обеспечивают возможность использования библиотек машинного обучения и анализа данных, таких как TensorFlow, Keras и Pandas [9, 10].

Цель настоящей работы состоит в разработке программной среды имитационного моделирования алгоритмов управления электроэнергией в гибких рас-

пределённых энергосетях на базе многоагентного подхода и рыночного механизма двойного аукциона. Среда реализована на Python с использованием фреймворка SPADE и протокола XMPP.

Структура системы

Рассматриваемая система включает четыре типа экономических агентов: агент-потребитель, агент-генератор (солнечный и ветровой), агент-накопитель и агент рынка. Каждый агент действует как независимый экономический субъект и принимает решения локально — исходя из собственных целей и доступной информации о прошлых значениях рыночной цены.

Агент-потребитель. Потребитель рассматривается как рациональный покупатель электроэнергии, максимизирующий функцию полезности за вычетом затрат. Применяется функция полезности квадратичного вида с убывающей предельной полезностью [11], имеющая следующий вид:

$$U(q) = aq - \frac{1}{2}bq^2 - Pq \quad (1)$$

где $U(q)$ — полезность от потребления q единиц энергии; a — коэффициент готовности платить; b — коэффициент ценовой чувствительности; P — текущая рыночная цена энергии; q — объём потребления энергии, кВт·ч. Из условия первого порядка выводится оптимальный объём потребления $q^* = (a - P) / b$ при $P < a$ и $q^* = 0$ в противном случае. Цена заявки отражает предельную готовность платить за дополнительную единицу энергии, объём заявки определяется максимальным запланированным потреблением. Временная зависимость потребления задаётся профилями с относительными коэффициентами, отражающими стандартные суточные колебания спроса [12–14].

Агент-генератор. Генераторы на основе возобновляемых источников энергии стремятся к максимизации прибыли. Целевая функция имеет вид:

$$\pi(q) = Pq - C(q) \quad (2)$$

где $\pi(q)$ — прибыль от производства q единиц энергии; P — рыночная цена продажи; $C(q)$ — функция издержек производства. Функция предельных издержек задаётся линейной зависимостью: к базовому значению добавляется член, пропорциональный отношению текущего объёма к установленной мощности генератора, с коэффициентом волатильности. Цена заявки на продажу устанавливается с наценкой над предельными издержками, обеспечивающей прибыльность.

Солнечная генерация моделируется как функция часа суток с учётом погодного коэффициента инсоляции

(принимая значения 1,0 при ясной погоде, 0,5 при облачной и 0,2 в дождливую) и коэффициента состояния оборудования. Ветровая генерация определяется аналогичным образом с коэффициентом силы ветра, принимающим максимальное значение в дождливые дни. Полученные зависимости воспроизводят типичные суточные профили генерации на основе возобновляемых источников.

Агент-накопитель. Система хранения энергии реализует стратегию межвременной оптимизации (ценового арбитража), заряжаясь при низких ценах и разряжаясь при высоких. Задача формулируется в виде:

$$\max \sum [P_t \cdot q_t^{\text{sell}} - P_t \cdot q_t^{\text{buy}}] \quad (3)$$

где суммирование ведётся по горизонту планирования; цена энергии и объёмы продажи и покупки в момент t рассматриваются как зависящие от времени переменные (цена измеряется в руб/кВт·ч). Задача решается при ограничениях на уровень заряда батареи, её ёмкость и мощности зарядки и разрядки. Для прогнозирования цены применяется скользящее среднее с линейной экстраполяцией тренда по заданному окну.

Агент рынка. Агент рынка реализует механизм двойного аукциона и выступает централизованным аукционистом. Для каждого временного шага имитации собираются анонимные заявки на покупку и продажу. Заявки на покупку сортируются по убыванию цены, формируя кривую совокупного спроса; заявки на продажу — по возрастанию цены, формируя кривую совокупного предложения. Совокупный спрос $D(P)$ определяется как суммарный объём заявок на покупку по цене не ниже P ; совокупное предложение $S(P)$ — как суммарный объём заявок на продажу по цене не выше P . Равновесная рыночная цена определяется на пересечении кривых спроса и предложения:

$$P_e \in [\max\{P_s(i) \mid S(P_s(i)) \leq D(P_s(i))\}, \min\{P_b(j) \mid D(P_b(j)) \leq S(P_b(j))\}] \quad (4)$$

Равновесный объём торгов для данного интервала определяется соотношением:

$$Q_e = \max_p \min\{D(P), S(P)\} \quad (5)$$

Заявки на покупку с ценой не ниже равновесной и заявки на продажу с ценой не выше равновесной признаются успешными, остальные отклоняются. Указанный механизм гарантирует расчёт по единой цене, определяемой совокупным спросом и предложением [15–20].

Управление имитацией выполняется классом-распорядителем, последовательно вызывающим метод расчёта у каждого активного агента. На каждом шаге каждый

агент анализирует текущее состояние рынка, рассчитывает оптимальную стратегию и формирует заявку. Собранный пул заявок передаётся агенту рынка, который применяет алгоритм определения равновесия и возвращает равновесную цену и распределение объёмов. Распорядитель обновляет глобальное состояние системы, присваивая текущей рыночной цене полученное равновесное значение, используемое агентами при принятии решений на следующем шаге.

Модель предполагает совершенную конкуренцию: ни один агент не обладает рыночной властью, отсутствуют комиссии, налоги и затраты на заключение сделок, агенты наблюдают текущую рыночную цену без задержек. Все агенты строго максимизируют свои целевые функции и не применяют стратегических манипуляций с заявками. Из рассмотрения в текущей версии модели исключены потери в линиях электропередач, ограничения пропускной способности, влияние топологии сети и реактивная мощность; их учёт предполагается при развитии модели.

Результаты

В ходе симуляции проверялась работоспособность разработанного алгоритма управления электроэнергией на базе механизма двойного аукциона. Моделирование охватывало период в один месяц. Алгоритм реали-

зует реактивное ценообразование на основе текущего баланса энергосистемы: при избытке генерации цена снижается, стимулируя потребление; при дефиците — растёт, ограничивая спрос.

Усреднённые суточные профили, представленные на Рисунке 1, показывают, что при преимущественно солнечной генерации в дневные часы (с 6:00 до 18:00) формируется избыток электроэнергии, и цена снижается до минимального уровня 0,5 руб/кВт·ч. В ночные часы при отсутствии солнечной генерации и сохранении потребления цена возрастает до 1,3 руб/кВт·ч. Эффективность алгоритма подтверждается отрицательной корреляцией между ценой и балансом энергосистемы.

Статистически значимая отрицательная корреляция, представленная на Рисунке 2, подтверждает корректность реакции механизма ценообразования на состояние энергосистемы. Для демонстрации работы алгоритма в более коротком временном масштабе получена детальная динамика за период трёх суток.

На Рисунке 3 прослеживается относительно стабильное потребление электроэнергии с характерными утренним и вечерним пиками, а также адаптивное изменение цены в ответ на отклонения между генерацией и потреблением. Агент-накопитель реализует стратегию ценового арбитража: за период моделирования зафик-

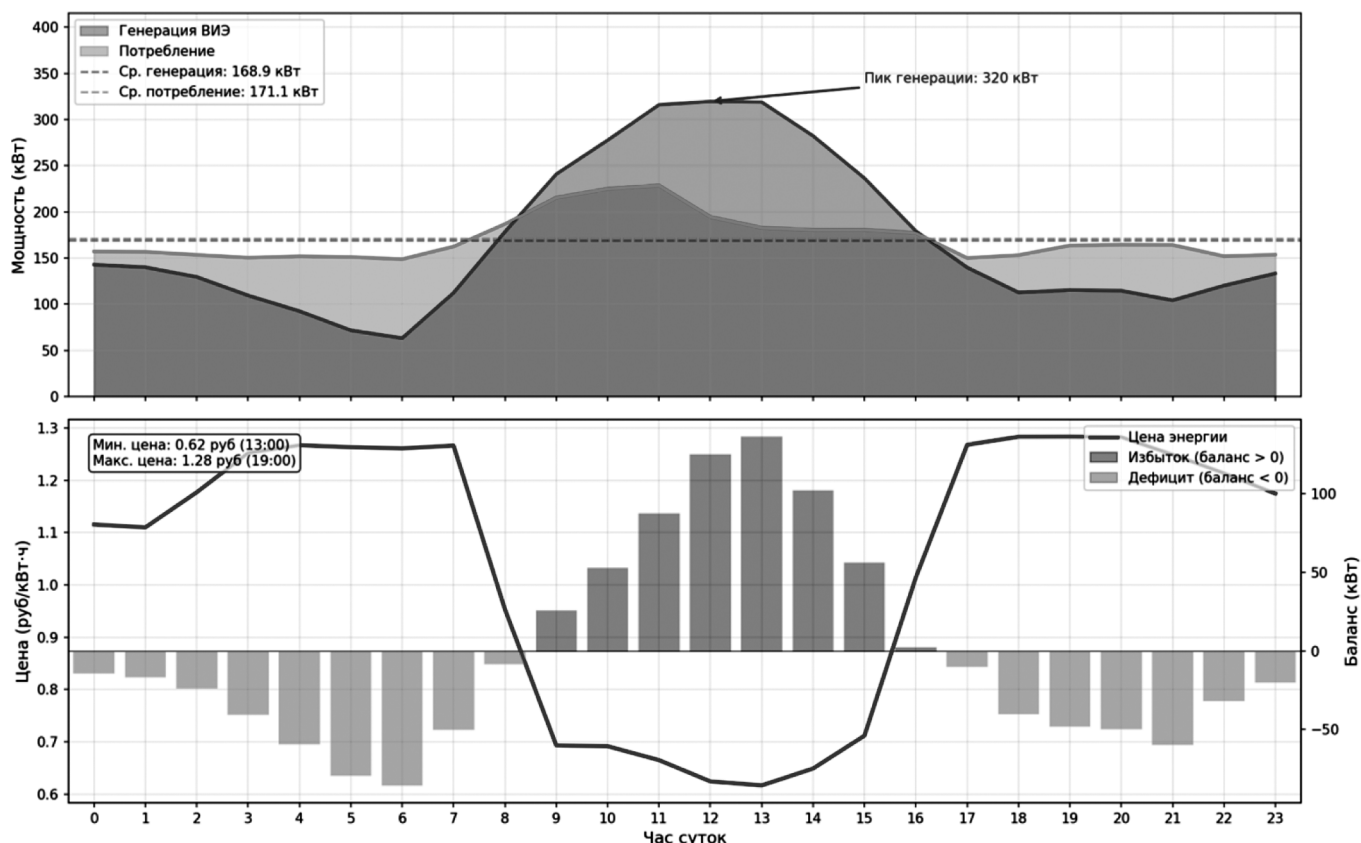


Рис. 1. Мощность генерации и баланс системы (усреднённые показатели)

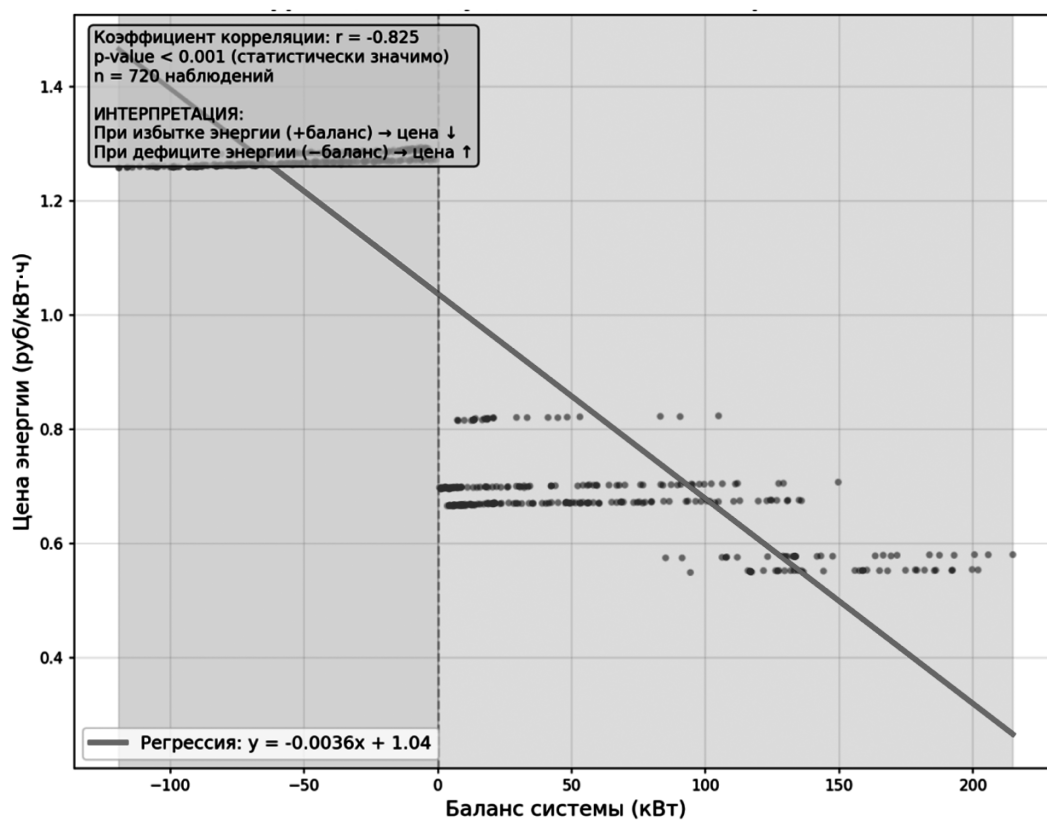


Рис. 2. Корреляция между ценой и балансом в системе

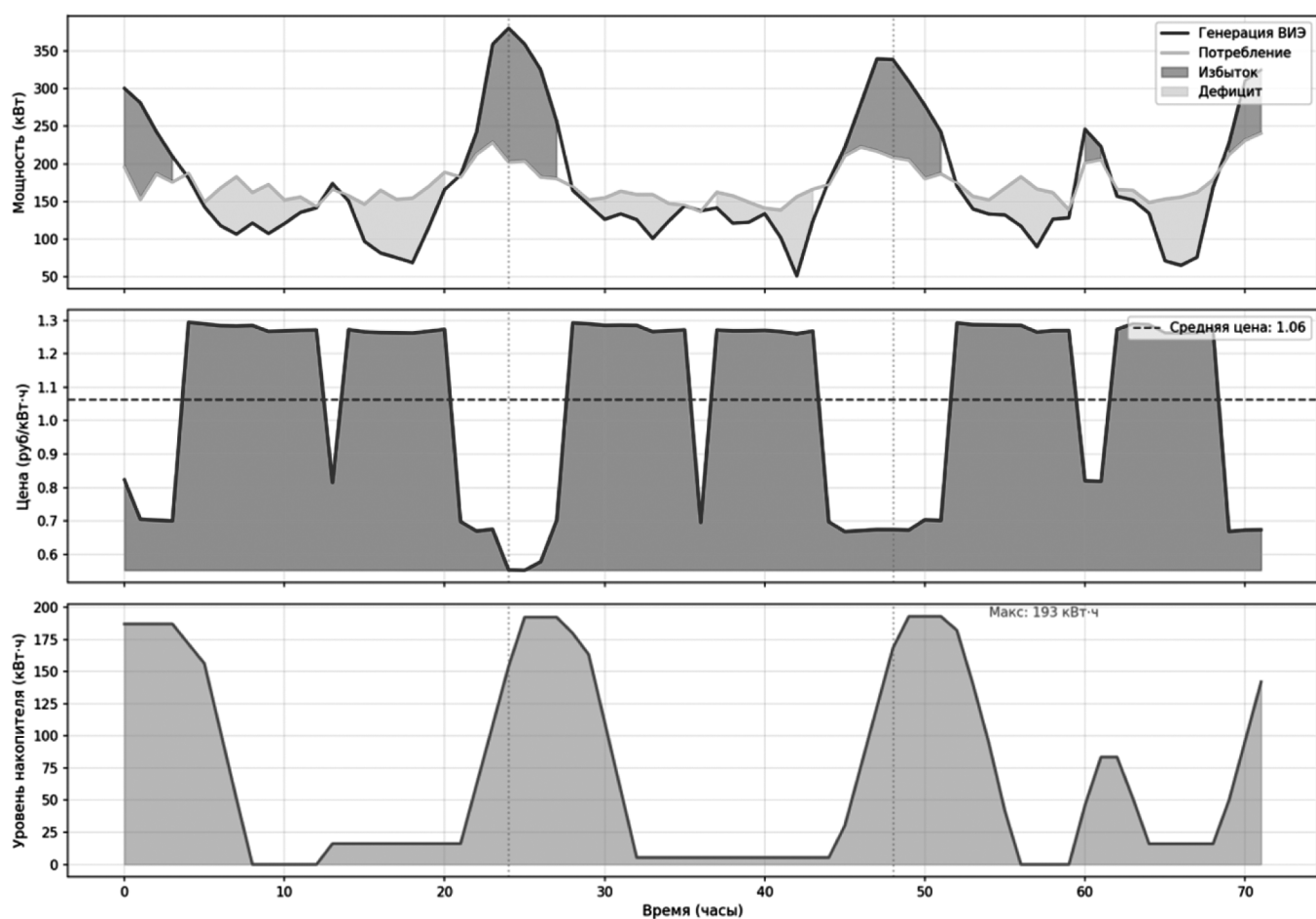


Рис. 3. Детальная динамика работы системы за трое суток

сировано 168 циклов зарядки при средней цене 0,7 руб/кВт·ч и 159 циклов разрядки при средней цене 1,2 руб/кВт·ч. Полученная разность цен подтверждает эффективность реализации стратегии и способствует сглаживанию пиковых нагрузок, что повышает общую стабильность системы.

Обсуждение

Полученные результаты согласуются с данными предшествующих исследований по рыночным механизмам координации в распределённых энергосетях [15, 18, 20]. Применение двойного аукциона обеспечивает формирование единой равновесной цены, прозрачной для всех участников, и устраняет необходимость согласования параметров напрямую между агентами. Децентрализованная организация взаимодействия через протокол XMPP сохраняет масштабируемость решения и позволяет добавлять новых агентов без перестройки архитектуры.

Ограничения рассмотренной модели связаны с допущениями о совершенной конкуренции и отсутствии физических ограничений сети. Дальнейшее развитие предполагает включение потерь в линиях электропере-

дач, тепловых пределов пропускной способности, влияния топологии и реактивной мощности, что приблизит поведение системы к реальным условиям эксплуатации. Отдельное перспективное направление — интеграция больших языковых моделей в логику принятия решений агентами: подобный подход расширяет пространство стратегий и позволяет учитывать сложные контекстные факторы, недоступные при формализации в виде замкнутых целевых функций.

Заключение

Разработана программная среда имитационного моделирования алгоритмов управления электроэнергией в микрогриде на базе рыночного механизма двойного аукциона и многоагентного подхода. Формализованы целевые функции для четырёх типов агентов — потребителей, солнечных и ветровых генераторов, накопителей и агента рынка. Описан механизм определения равновесной цены и объёма торгов на каждом шаге. Результаты подтвердили работоспособность алгоритма: зафиксирована статистически значимая отрицательная корреляция между ценой и балансом системы, продемонстрирована эффективность стратегии ценового арбитража накопителей.

ЛИТЕРАТУРА

1. Arora A. et al. Next generation of multi-agent driven smart city applications and research paradigms //IEEE Open Journal of the Communications Society. — 2023. — Т. 4. — С. 2104–2121.
2. Palanca J. et al. A flexible agent architecture in SPADE //International Conference on Practical Applications of Agents and Multi-Agent Systems. — Cham: Springer International Publishing, 2022. — С. 320–331.
3. Matagne M. M. Developing a framework for SPADE agents management that implement consensus learning algorithms. — 2023. — 87 с.
4. Laseter R. MicroGrids // Proceedings of the IEEE Power Engineering Society Winter Meeting. 2002. Т. 1. — С. 305–308.
5. Altin N., Eyimaya S. E., Nasiri A. Multi-agent-based controller for microgrids: An overview and case study //Energies. — 2023. — Т. 16. — №. 5. — С. 2445.
6. Hamidi M., Raihani A., Bouattane O. Sustainable intelligent energy management system for microgrid using multi-agent systems: A case study //Sustainability. — 2023. — Т. 15. — №. 16. — С. 12546.
7. Uddin M. et al. Microgrids: A review, outstanding issues and future trends //Energy Strategy Reviews. — 2023. — Т. 49. — С. 101127.
8. Bellifemine F., Poggi A., Rimassa G. Developing multi-agent systems with JADE // International Workshop on Agent Theories, Architectures, and Languages. Berlin, Heidelberg: Springer, 2000. — С. 89–103.
9. Ferraz Junior N. et al. Performance evaluation of publish-subscribe systems in IoT using energy-efficient and context-aware secure messages //Journal of Cloud Computing. — 2022. — Т. 11. — №. 1. — С. 6.
10. Saint-Andre P., Smith K., Tronçon R. XMPP: The Definitive Guide. Sebastopol: O'Reilly Media, 2009. 310 с.
11. Gao Y. et al. Intelligent building energy management considering uncertainties and variation of day-ahead and real-time electricity prices //IEEE Internet of Things Journal. — 2024. — Т. 11. — №. 23. — С. 37785–37798.
12. Zhang L. et al. Bi-level stochastic real-time pricing model in multi-energy generation system: A reinforcement learning approach //Energy. — 2022. — Т. 239. — С. 121926.
13. Malik S. et al. Comparative double auction approach for peer-to-peer energy trading on multiple microgrids //Smart Grids and Sustainable Energy. — 2023. — Т. 8. — №. 4. — С. 21.
14. Izmirliglu Y. et al. A survey of multi-agent systems for smart grids //Energies. — 2024. — Т. 17. — №. 15. — С. 3620.
15. Arévalo P. et al. Systematic review of hierarchical and multi-agent optimization strategies for P2P energy management and electric machines in microgrids //Applied Sciences. — 2025. — Т. 15. — №. 9. — С. 4817.
16. Azim M. I. et al. Peer-to-peer kilowatt and negawatt trading: A review of challenges and recent advances in distribution networks //Renewable and Sustainable Energy Reviews. — 2022. — Т. 169. — С. 112908.
17. Huang W., Li H. Game theory applications in the electricity market and renewable energy trading: A critical survey //Frontiers in Energy Research. — 2022. — Т. 10. — С. 1009217.
18. Doan H. T., Nam H., Kim D. Optimal peer-to-peer energy trading under load uncertainty incorporating carbon emission and transaction cost for grid-connected prosumers //IEEE Access. — 2022. — Т. 10. — С. 106202–106216.
19. Zhang R. et al. Distributed peer to peer transaction framework for heterogeneous energy trading in urban energy internet-integrated micro-energy grid systems with diverse preference aware //International Journal of Electrical Power & Energy Systems. — 2025. — Т. 173. — С. 111332.
20. Gao G., Wen Y., Tao D. Distributed energy trading and scheduling among microgrids via multiagent reinforcement learning //IEEE transactions on neural networks and learning systems. — 2022. — Т. 34. — №. 12. — С. 10638–10652.

© Цай Цин (caiq041@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»

ОБЕСПЕЧЕНИЕ КИБЕРБЕЗОПАСНОСТИ НА ПРОМЫШЛЕННЫХ ПРЕДПРИЯТИЯХ С ПОМОЩЬЮ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

ENSURING CYBERSECURITY IN INDUSTRIAL ENTERPRISES USING ARTIFICIAL INTELLIGENCE

**I. Cherenkova
O. Kishkinova
I. Kutlikova**

Summary. The use of artificial intelligence to ensure the cybersecurity of industrial enterprises is one of the most promising areas for the development of critical infrastructure protection systems. The article is devoted to the analysis of the use of artificial intelligence technologies to ensure the cybersecurity of industrial enterprises in the context of digital transformation. The author systematizes modern approaches to the use of neural network architectures for detecting cyber attacks on industrial automation systems, including examples of real-world developments by research teams. It has been revealed that the introduction of AI solutions in industrial cybersecurity is hampered by a complex of technological barriers (outdated equipment, protocol incompatibility), organizational (uncertainty of effects, lack of cases), personnel (lack of interdisciplinary specialists) and legal (unallocated responsibility). The main unresolved problem remains the «explainability» of neural network model solutions, which is most relevant in the context of multidimensional time series analysis of industrial data. Specific recommendations have been proposed to overcome them. The author concludes that solving the identified problems and overcoming barriers requires a systematic approach, including infrastructure modernization, the development of educational programs, the creation of a regulatory framework and the formation of industry standards for the safe use of AI in critical processes. The research results can be used in the development of digital transformation strategies for cybersecurity at industrial facilities.

Keywords: artificial intelligence, AI, cybersecurity, anomaly detection, machine learning, neural networks.

Черенкова Ирина Анатольевна

Старший преподаватель, Московская государственная академия ветеринарной медицины и биотехнологии имени К.И. Скрябина
9738296@mail.ru

Кишкинова Ольга Алексеевна

Старший преподаватель, Московская государственная академия ветеринарной медицины и биотехнологии имени К.И. Скрябина
olga.19672015@yandex.ru

Кутликова Ирина Вениаминовна

Старший преподаватель, Московская государственная академия ветеринарной медицины и биотехнологии имени К.И. Скрябина
iri68na68@mail.ru

Аннотация. Применение искусственного интеллекта для обеспечения кибербезопасности промышленных предприятий представляет собой одно из наиболее перспективных направлений развития систем защиты критической инфраструктуры. Статья посвящена анализу применения технологий искусственного интеллекта для обеспечения кибербезопасности промышленных предприятий в условиях цифровой трансформации. Автором систематизированы современные подходы к использованию нейросетевых архитектур для детектирования кибератак на системы промышленной автоматизации, включая примеры реальных разработок научных коллективов. Выявлено, что внедрение ИИ-решений в промышленной кибербезопасности сдерживается комплексом барьеров технологического (устаревшее оборудование, несовместимость протоколов), организационного (неопределённость эффектов, нехватка кейсов), кадрового (дефицит междисциплинарных специалистов) и правового (нераспределённая ответственность) характера. Главной нерешённой проблемой остаётся «объяснимость» решений нейросетевых моделей, что наиболее актуально в контексте анализа многомерных временных рядов промышленных данных. Для их преодоления были предложены конкретные рекомендации. Автор приходит к выводу, что решение выявленных проблем и преодоление барьеров требует системного подхода, включающего модернизацию инфраструктуры, развитие образовательных программ, создание нормативно-правовой базы и формирование отраслевых стандартов безопасного применения ИИ в критических процессах. Результаты исследования могут быть использованы при разработке стратегий цифровой трансформации кибербезопасности на промышленных объектах.

Ключевые слова: искусственный интеллект, ИИ, кибербезопасность, обнаружение аномалий, машинное обучение, нейронные сети.

Введение

Цифровая трансформация промышленного сектора, выражающаяся во внедрении технологий промышленного интернета вещей, интеллектуальных систем управления и предиктивной аналитики, значительно изменила ландшафт угроз для промышленных предприятий. Раньше промышленные сети, которые управляют оборудованием, работали отдельно от офисных компьютерных сетей — между ними был физический разрыв, так называемый «воздушный зазор». Это делалось для безопасности. Но сегодня ради повышения эффективности производства эти сети всё чаще соединяют друг с другом. Такое сближение информационных и операционных технологий создаёт новые векторы атак: злоумышленники получают возможность проникать в производственные сети через корпоративный сегмент, что делает традиционные методы обнаружения угроз недостаточно эффективными. В этих условиях технологии искусственного интеллекта рассматриваются как один из наиболее перспективных инструментов повышения уровня защищённости промышленных объектов. Однако внедрение ИИ-решений в промышленной кибербезопасности сопряжено с рядом специфических вызовов, отличающих этот процесс от применения аналогичных технологий в корпоративных информационных системах.

Цель исследования состоит в системном анализе особенностей применения искусственного интеллекта для обеспечения кибербезопасности промышленных предприятий, включая классификацию используемых методов, обобщение практического опыта российских разработчиков и исследователей, а также идентификацию барьеров и направлений их преодоления.

Материалы и методы исследования

Теоретико-методологическую основу исследования составили научные труды и публикации российских и зарубежных авторов в области промышленной кибербезопасности, обнаружения аномалий на основе машинного обучения, а также отраслевые исследования по внедрению искусственного интеллекта на промышленных предприятиях. Методология включала системный анализ, сравнительный анализ, а также проблемно-ориентированный подход к идентификации барьеров внедрения.

Результаты и обсуждения

Применение искусственного интеллекта в промышленной кибербезопасности охватывает широкий спектр задач: от анализа сетевого трафика и поведения пользователей до мониторинга целостности технологических параметров и прогнозирования атак. Главное преимущество ИИ-решений заключается в их способности об-

учаться на основе нормативных паттернов функционирования промышленных систем и выявлять аномалии, которые не могут быть обнаружены традиционными методами. Например, машинное обучение успешно применяется для обнаружения компрометации цепочек поставок, эксплуатации уязвимостей, целевых атак социальной инженерии и сложных многоэтапных инцидентов.

В российской науке и практике накоплен значительный опыт создания систем кибербезопасности на основе искусственного интеллекта для промышленного сектора. В исследовании Д.Б. Краснослободцевой и А.В. Юдина представлен комплексный сравнительный анализ эффективности современных нейросетевых архитектур для противодействия целевым атакам социальной инженерии на промышленные системы. Эксперименты на наборе данных, содержащем более 651 тысячи URL-адресов, показали, что гибридные архитектуры значительно превосходят отдельные модели. Комбинация свёрточной и рекуррентной нейронных сетей достигла максимальной точности 92,29 %, а гибридный CNN с сетями Колмогорова-Арнольда показала результат 92 %. Капсульные нейронные сети наиболее эффективно идентифицировали безопасные ресурсы с точностью 98,60 %, тогда как гибридный CNN+LSTM лучше всего обнаруживал фишинговые атаки (72,76 %). Авторы установили корреляцию между типом нейросетевой архитектуры и характером возможной киберугрозы, что создаёт методологическую основу для разработки адаптивных систем безопасности нового поколения для промышленной инфраструктуры [3].

Исследователи Санкт-Петербургского Федерального исследовательского центра РАН создали нейросетевой фреймворк ForecaState для выявления киберугроз в системах промышленного интернета вещей. Эксперименты на большом объёме данных промышленных систем, включая систему очистки воды и сеть электрических трансформаторов, продемонстрировали снижение числа ошибок при выявлении угроз на 30 % по сравнению с существующими аналогами. По словам разработчиков, модели глубокого обучения нейросетей хорошо подходят для обработки больших объёмов данных, характерных для технологий IIoT [6].

Еще одну модификацию предложили учёные Южно-Уральского государственного университета, разработавшие метод мониторинга промышленных систем на основе нейронной сети Кохонена. Их модель анализирует показания всех датчиков в штатном режиме, создавая карту типичных состояний, а затем переходит в режим постоянного мониторинга, сравнивая поступающие данные с эталонной картиной. При значительном отклонении от нормы система мгновенно сигнализирует об аномалии. Тесты подтвердили эффективность моде-

ли: система правильно классифицировала 94 % данных, а время первоначального обучения на штатных параметрах заняло около 3,5 минут [9].

Одной из самых серьёзных проблем, ограничивающих практическое применение ИИ-моделей в промышленной кибербезопасности, является их «чёрный ящик» — невозможность для человека-оператора понять, на каком основании нейросеть приняла то или иное решение. Эта проблема особенно остра на промышленных объектах критической инфраструктуры, где от своевременного и точного определения источника аномалии зависят безопасность людей и непрерывность производственных процессов.

Исследователи СПб ФИЦ РАН столкнулись с подобной проблемой, разработав метод обнаружения кибератак на критически важные объекты на основе преобразования потоков данных с датчиков в матричные «изображения» с последующим анализом свёрточными нейронными сетями. Разработанный ими ИИ научился находить аномалии, продемонстрировав точность до 99 %. Однако при попытке объяснить, на основании чего модель принимала решения, возникли серьёзные трудности. Три современных метода объяснимости — SHAP, Grad-CAM и Guided Grad-CAM — показали непригодные для практики результаты: точность локализации датчика-виновника составила от 0,2 до 60 % в лучших сценариях и менее 0,01 % в худших. Причина кроется в природе промышленных данных — это многомерные временные ряды, где показания взаимосвязаны как в пространстве, так и во времени, тогда как испытанные методы не учитывают временные зависимости. Дальнейшие исследования авторы планируют направить в сторону графовых нейронных сетей и разработки принципиально новых методов объяснимости, учитывающих временную динамику [6].

Актуальность разработки «объяснимого» искусственного интеллекта для промышленной кибербезопасности подтверждается и зарубежными исследованиями. Авторы Н. Skrodelis, A. Romānovs, A. Liekniņš выдвигают гипотезу о том, что гибридная система обнаружения вторжений, объединяющая сигнатурный движок с детектором аномалий на основе машинного обучения и дополненная объяснениями для оператора через технологии XAI (например, SHAP-значения), может существенно повысить операционную полезность систем защиты промышленных сетей [10].

Вторым важнейшим направлением применения искусственного интеллекта в промышленной кибербезопасности является автоматизация процессов реагирования на инциденты. Традиционные центры мониторинга безопасности генерируют десятки и сотни подозрений на инциденты ежедневно, что делает их ручной анализ практически невозможным. По оценкам аналитиков,

среднее время обнаружения кибератаки составляет около 200 дней, а полное устранение инцидента может занимать более 70 дней. Согласно прогнозам Gartner, уже к 2027 году более 40 % операций центров мониторинга безопасности будут выполняться автоматически. Главным направлением становится использование искусственного интеллекта для автоматического подключения новых источников данных в SIEM-системы и быстрой разработки корреляционной логики, что позволяет выявлять угрозы ещё до того, как они приведут к значительным инцидентам [2]. На российском рынке появляются соответствующие решения. Так, группа компаний «Солар» представила платформу Solar SIEM, объединяющую функции SIEM и SOAR с встроенным нейроромощником на базе ИИ. Ассистент помогает аналитикам формировать правила корреляции, писать запросы к данным и создавать сценарии реагирования [4].

Эксперты прогнозируют [2], что к 2030 году классические SOC трансформируются в интеллектуальные центры безопасности, объединяющие аналитику угроз, автоматическое реагирование и прогнозирование атак. Эти smart SOC будут обеспечивать единую видимость инцидентов как в корпоративных ИТ-системах, так и в промышленной инфраструктуре, что особенно важно в условиях конвергенции ИТ- и ОТ-сред.

Несмотря на очевидные преимущества, массовое внедрение ИИ-решений в промышленной кибербезопасности сталкивается с серьёзными барьерами, которые можно классифицировать на технологические, организационные, правовые и кадровые.

Технологические барьеры связаны прежде всего с состоянием парка промышленного оборудования. Значительная часть промышленных контроллеров имеет возраст более десяти лет, их вычислительные мощности недостаточны для выполнения ИИ-моделей на борту. Несовместимость коммуникационных протоколов с современными системами управления создаёт дополнительные сложности.

Организационные барьеры включают сопротивление персонала и неопределённость экономических эффектов. Согласно совместному исследованию ФГАУ «Цифровые промышленные технологии» и ИТ-компании «K2Tech», 61 % промышленных предприятий назвали плохо прогнозируемые эффекты главным препятствием при внедрении ИИ. Более 53 % респондентов указали на риски кибербезопасности как на барьер, более 46 % — на нехватку реальных эффективных кейсов, а более 38 % — на отсутствие квалифицированных кадров. Полностью готовая к внедрению ИИ-решений ИТ-инфраструктура оказалась лишь у 7 % предприятий [1]. Кадровый дефицит проявляется в нехватке специалистов, одновременно владеющих компетенциями в области промышленных технологий, кибербезопасно-

сти и машинного обучения. Правовой барьер остаётся одним из самых труднопреодолимых. В критически важных технологических процессах окончательное решение пока остаётся за человеком. Главная причина — нерешённый вопрос об ответственности в случае аварии, вызванной решением алгоритма [5].

Нельзя обойти вниманием и риски, связанные с самими системами искусственного интеллекта как объектами атак. Здесь важно отметить риск внедрения злоумышленниками искажённой информации в обучающие выборки, что снижает точность моделей и может привести к сбоям оборудования. По данным, представленным на форуме «Киберустойчивая Арктика 2025», убытки от такого инцидента могут достигать 15 миллионов рублей, а вероятность аварий возрастает на 50 % [1].

В таблице 1 систематизированы основные проблемы внедрения ИИ в промышленную кибербезопасность и предложены пути их решения.

Следовательно, преодоление выявленных барьеров требует системного подхода, включающего модернизацию инфраструктуры, развитие образовательных программ, создание нормативно-правовой базы и формирование отраслевых стандартов безопасного применения ИИ в критических процессах.

Выводы

В данной статье проведен анализ применения технологий искусственного интеллекта для обеспечения кибербезопасности промышленных предприятий в условиях цифровой трансформации. Автором систематизированы современные подходы к использованию нейросетевых архитектур для детектирования кибератак на системы промышленной автоматизации, включая примеры реальных разработок научных коллективов. Выявлено, что внедрение ИИ-решений в промышленной кибербезопасности сдерживается комплексом барьеров технологического (устаревшее оборудование, несовместимость протоколов), организационного (неопределённость эффектов, нехватка кейсов), кадрового (дефицит междисциплинарных специалистов) и правового (нераспределённая ответственность) характера. Главной нерешённой проблемой остаётся «объяснимость» решений нейросетевых моделей, что наиболее актуально в контексте анализа многомерных временных рядов промышленных данных. Для их преодоления были предложены конкретные рекомендации: модернизация инфраструктуры, развитие образовательных программ, создание нормативно-правовой базы и формирование отраслевых стандартов безопасного применения ИИ в критических процессах. Результаты исследования могут быть использованы при разработке стратегий цифровой трансформации кибербезопасности на промышленных объектах.

Таблица 1.

Проблемы и их решение при внедрении ИИ-технологий кибербезопасности на промышленных предприятиях

Проблема	Решение
Необъяснимость решений нейросетей («чёрный ящик»), невозможность для оператора верифицировать выводы системы	Внедрение технологий объяснимого искусственного интеллекта (XAI), адаптированных к многомерным временным рядам промышленных данных; разработка графовых и гибридных моделей с сохранением временной динамики
Дефицит маркированных и репрезентативных наборов данных для обучения моделей на реальных промышленных атаках	Использование синтетических данных, генерируемых на основе реальных технологических процессов; создание открытых отраслевых датасетов и платформ для обмена анонимизированными данными об инцидентах
Высокая стоимость модернизации устаревшего оборудования и несовместимость протоколов	Применение модульной архитектуры с поэтапной модернизацией без остановки производства; использование адаптеров и шлюзов для интеграции разнородных протоколов
Отсутствие специалистов с междисциплинарными компетенциями (промышленные технологии + кибербезопасность + машинное обучение)	Разработка и внедрение специализированных образовательных программ; создание центров компетенций на базе ведущих университетов и отраслевых институтов
Правовая неопределённость ответственности за последствия решений, принятых ИИ в автоматическом режиме	Разработка нормативно-правовой базы, определяющей границы автономности ИИ-систем и механизмы распределения ответственности между разработчиком, владельцем и оператором
Уязвимость самих ИИ-моделей к атакам (отравление данных, подмена входных параметров)	Внедрение практик безопасной разработки ИИ (Secure AI), верификация цепочки поставок данных, мониторинг целостности обучающих выборок с использованием криптографических методов
Низкое доверие руководства предприятий к решениям ИИ в критических процессах	Проведение пилотных проектов на некритичных участках с последующей масштабируемой апробацией; создание реестров успешных кейсов и методических рекомендаций
Высокая стоимость полной замены существующих систем мониторинга	Предпочтение модульных гибридных архитектур, позволяющих постепенно наращивать функционал без «замены под ключ»; использование облачных сервисов ИИ-аналитики

Источник: составлено автором по данным [1, 3, 5, 7–11]

ЛИТЕРАТУРА

1. Иванов, И. Неизвестность пугает. Более половины компаний назвали плохо прогнозируемые эффекты главным препятствием при внедрении ИИ // Comnews. — URL: <https://www.comnews.ru/content/240291/2025-07-22/2025-w30/1008/neizvestnost-pugaet-bolee-poloviny-kompaniy-nazvali-plokho-prognoziруемые-эффекты-главным-препятствием-pri-vnedrenii-ii> (дата обращения: 20.05.2026).
2. К 2030 году центры кибербезопасности станут «умными» // ИТ-клуб. — URL: https://www.it-world.ru/view/233756.html?PAGEN_1=3&IBLOCK_CODE=view (дата обращения: 20.05.2026).
3. Краснослободцева, Д.Б., Юдин А.В. Анализ эффективности нейросетевых архитектур для защиты промышленных систем от целевых атак социальной инженерии // Computational nanotechnology. — 2025. — Т. 12, № 5. — С. 95–109. — DOI: 10.33693/2313-223X-2025-12-5-95-109.
4. Лященко, И. «Солар» запускает умный SIEM с ИИ-ассистентом и модульной архитектурой // Журнал о науке и технологиях «Компьютерра». — URL: <https://www.computerra.ru/326946/solar-zapuskayet-umnyj-siem-s-ii-assistentom-i-modulnoj-arhitekturoj/> (дата обращения: 20.05.2026).
5. Польская, С.И. Искусственный интеллект в кибербезопасности: этические аспекты защиты данных, предотвращения утечек и ответственности исследователя / С.И. Польская, Л.В. Польская // Проблемы информационной безопасности социально-экономических систем : Труды XII Международной научно-практической конференции, Гурзуф, 19–21 февраля 2026 года. — Симферополь: ИП Зуева Т.В., 2026. — С. 242–244.
6. Разработан цифровой инструмент для защиты от киберугроз на промышленных предприятиях // Российская академия наук (РАН). — URL: https://new.ras.ru/press-center/razrobotan-tsifrovoy-instrument-dlya-zashchity-ot-kiberugroz-na-promyshlennykh-predpriyatiyakh/?sphrase_id=18321253 (дата обращения: 20.05.2026).
7. Создан ИИ для поиска кибератак на промышленные системы, но возникла неожиданная проблема // Газета «Поиск». — URL: <https://poisknews.ru/grants/sozdan-ii-dlya-poiska-kiberatak-na-promyshlennye-sistemy-no-voznikla-neozhidannaya-problema/> (дата обращения: 20.05.2026).
8. Степашкина, В.П. Предсказание промышленных кибератак с использованием нормализующих потоков / В.П. Степашкина, М.И. Гушин // Доклады Российской академии наук. Математика, информатика, процессы управления. — 2024. — Т. 520, № S2. — С. 107–115. — DOI 10.31857/S2686954324700425.
9. Ученые ЮрГУ разработали нейросетевой метод защиты от кибератак // Журнал о науке и технологиях «Компьютерра». — URL: <https://www.computerra.ru/325708/uchenye-yurgu-razrabotali-nejrosetevoj-metod-zashchity-ot-kiberatak/> (дата обращения: 20.05.2026).
10. Skrodelis, H., Romānovs A., Liekņiņš A. Explainable SCADA Hybrid IDS: Gaps and Hypothesis. In: 2025 66th International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS 2025): Proceedings, Latvia, Riga, 9–10 October, 2025. IEEE, 2025. — pp. 11236701.
11. Zou, B., Zhang K.j., Yu X.Y. et al. An Anomaly Detection Method for Industrial System Cybersecurity Based on GGL-WAVE-CNN // International Journal of Computational Intelligence Systems. — 2025. — №18(1). — pp. 177. <https://doi.org/10.1007/s44196-025-00832-5>.

© Черенкова Ирина Анатольевна (9738296@mail.ru); Кишкинова Ольга Алексеевна (olga.19672015@yandex.ru);

Кутликова Ирина Вениаминовна (iri68na68@mail.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

АНАЛИЗ АРХИТЕКТУР И ТЕХНОЛОГИЙ ИНТЕРНЕТА ВЕЩЕЙ В РАСПРЕДЕЛЕННЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМАХ

ANALYSIS OF ARCHITECTURES AND TECHNOLOGIES OF THE INTERNET OF THINGS IN DISTRIBUTED COMPUTING SYSTEMS

**V. Cherepenin
D. Galushkina
S. Vorobyov**

Summary. This paper examines modern architectural approaches and technological solutions in the field of the Internet of Things (IoT). The main architectures of IoT systems are analyzed, including centralized, decentralized, and hybrid models. Particular attention is given to the multi-layer structure of IoT, which includes the sensing layer, network layer, data processing layer, and application layer. Key technologies for data transmission and processing are considered, such as wireless communication networks, edge and fog computing, cloud platforms, and artificial intelligence methods. It is shown that the integration of distributed computing architectures makes it possible to improve scalability, reliability, and efficiency of data processing in IoT environments.

Keywords: Internet of Things, distributed systems, Cloud-Fog-Edge architecture, data processing, network technologies, intelligent systems, digital services.

Черепенин Валентин Анатольевич

Аспирант, Южно-Российский государственный
политехнический университет (НПИ)
имени М.И. Платова (г. Новочеркасск)
blankzer@yandex.ru

Галушкина Дарья Владимировна

Аспирант, Южно-Российский государственный
политехнический университет (НПИ)
имени М.И. Платова (г. Новочеркасск)
dashagaluschkina@yandex.ru

Воробьев Сергей Петрович

кандидат технических наук, доцент, Южно-Российский
государственный политехнический университет (НПИ)
имени М.И. Платова (г. Новочеркасск)
vsp19999@yandex.ru

Аннотация. В статье рассматриваются современные архитектурные подходы и технологические решения в области Интернета вещей (IoT). Анализируются основные архитектуры IoT-систем, включая централизованные, децентрализованные и гибридные модели. Особое внимание уделяется многоуровневой структуре IoT, включающей сенсорный, сетевой, вычислительный и прикладной уровни. Рассматриваются ключевые технологии передачи и обработки данных, такие как беспроводные сети, edge- и fog-вычисления, облачные платформы и методы искусственного интеллекта. Показано, что интеграция распределённых вычислительных архитектур позволяет повысить масштабируемость, надёжность и эффективность обработки данных в IoT-среде.

Ключевые слова: Интернет вещей, распределенные системы, Cloud-Fog-Edge архитектура, обработка данных, сетевые технологии, интеллектуальные системы, цифровые сервисы.

Введение

Современное развитие цифровых технологий сопровождается стремительным ростом количества подключённых устройств, объединённых в инфраструктуру Интернета вещей (Internet of Things, IoT). IoT представляет собой сложную экосистему взаимосвязанных устройств, сенсоров и программных платформ, обеспечивающих сбор, передачу и анализ данных из физической среды.

В последние годы применение IoT активно расширяется в различных областях, включая промышленность, транспорт, медицину, энергетику и системы «умного города». Рост объёмов данных и увеличение количества устройств требуют создания эффективных архитектур-

ных решений, обеспечивающих масштабируемость, безопасность и устойчивость работы системы.

Ключевым направлением развития IoT является переход от традиционных централизованных архитектур к распределённым моделям обработки данных. В таких системах вычислительные ресурсы распределяются между облачными платформами, туманными узлами и периферийными устройствами, что позволяет уменьшить задержки и повысить эффективность обработки информации [1].

Экосистема IoT включает широкий спектр технологий, стандартов и протоколов взаимодействия устройств. Их анализ позволяет определить наиболее эффективные архитектурные подходы для построения

современных цифровых систем. В рамках исследования рассматриваются основные архитектурные модели IoT, уровни организации инфраструктуры и современные технологические тенденции, влияющие на развитие распределённых вычислительных систем.

Архитектурные подходы в системах Интернета вещей

Развитие Интернета вещей связано с формированием сложных распределённых инфраструктур, обеспечивающих взаимодействие большого количества устройств, сенсоров и программных сервисов. В зависимости от принципов организации обработки данных и управления системой можно выделить несколько основных архитектурных подходов: централизованные, децентрализованные и гибридные архитектуры IoT.

В зависимости от принципов организации обработки данных и управления системой [2] можно выделить несколько основных архитектурных подходов, показанных на рисунке 1: централизованные, распределённые и децентрализованные архитектуры IoT.

Централизованные архитектуры предполагают наличие единого вычислительного центра, в котором выполняется основная обработка и хранение данных. В такой модели все устройства Интернета вещей передают информацию на центральный сервер или облачную платформу, где осуществляется её анализ и принятие решений.

Преимущества централизованных архитектур являются простота управления системой, высокая вычислительная мощность центральных серверов и возможность использования масштабируемых облачных платформ. Кроме того, централизованная модель облегчает управление безопасностью и хранением данных [3].

Однако по мере роста количества устройств и объёма передаваемых данных централизованные системы сталкиваются с рядом ограничений. К ним относятся увеличение сетевых задержек, перегрузка каналов связи и высокая нагрузка на облачные ресурсы. В задачах, требующих обработки данных в реальном времени, такие ограничения могут существенно снижать эффективность работы системы.

Децентрализованные архитектуры предполагают распределение вычислительных ресурсов между множеством узлов сети. В таких системах обработка данных может выполняться непосредственно на устройствах или на локальных вычислительных узлах.

Основным преимуществом децентрализованного подхода является повышение устойчивости системы к отказам и снижению нагрузки на центральные серверы. Распределение вычислений позволяет уменьшить задержки передачи данных и обеспечить более быструю реакцию системы на изменения в окружающей среде [4].

Кроме того, децентрализованные архитектуры способствуют повышению уровня масштабируемости, поскольку новые устройства могут подключаться к системе без существенного изменения её структуры. Тем не менее управление такими системами требует более сложных алгоритмов координации и синхронизации данных между узлами сети.

Наиболее перспективным направлением развития IoT является использование гибридных архитектур, сочетающих элементы централизованных и децентрализованных систем. В таких моделях часть вычислений выполняется на периферийных устройствах, часть — на промежуточных узлах сети, а наиболее ресурсоёмкие задачи обрабатываются в облачных центрах обработки данных [5].

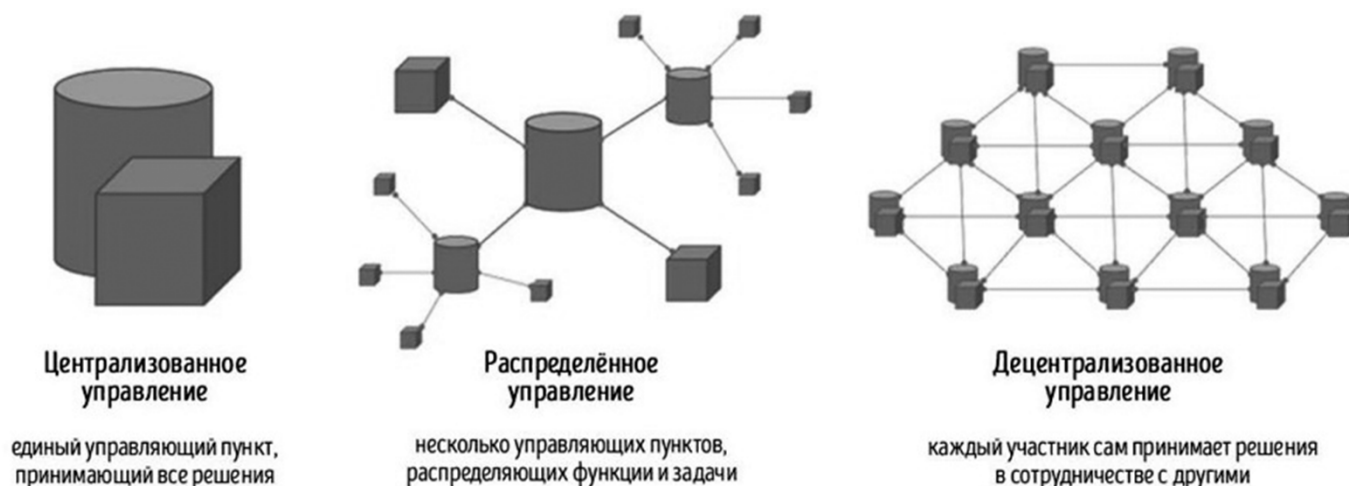


Рис. 1. Архитектурные подходы IoT

Подобная архитектура получила широкое распространение в рамках концепции Cloud-Fog-Edge computing. В этой модели периферийный уровень (edge) обеспечивает обработку данных непосредственно на устройствах или рядом с источником их генерации. Fog-уровень выполняет функции локальной агрегации, предварительной обработки и маршрутизации данных. Облачный уровень отвечает за долгосрочное хранение информации, сложную аналитику и обучение интеллектуальных моделей.

Интеграция различных уровней вычислений позволяет значительно снизить задержки передачи данных и повысить эффективность использования сетевых ресурсов. Кроме того, гибридные архитектуры обеспечивают высокую масштабируемость и устойчивость IoT-систем при увеличении числа подключённых устройств.

Использование распределённых моделей обработки данных становится особенно актуальным в условиях стремительного роста количества IoT-устройств и увеличения объёмов генерируемой информации. Такие архитектуры обеспечивают более гибкое распределение вычислительной нагрузки и повышают надёжность функционирования цифровых сервисов.

Сравнительный анализ архитектур IoT в распределённых вычислительных системах

Современные системы Интернета вещей могут строиться на основе различных архитектурных моделей, каждая из которых обладает своими преимуществами и ограничениями. Выбор архитектуры определяется требованиями к масштабируемости, задержкам обработки данных, уровню безопасности и характеру используемых приложений. В условиях роста объёмов данных и увеличения количества подключённых устройств особое значение приобретает возможность эффективного распределения вычислительных ресурсов.

Централизованные архитектуры характеризуются наличием единого центра управления и обработки данных. Основным преимуществом данного подхода является упрощённое администрирование и высокая вычислительная мощность центральных серверов. При увеличении количества устройств и объёмов передаваемой информации такие системы сталкиваются с проблемами масштабируемости, перегрузкой каналов связи и увеличением задержек передачи данных.

Децентрализованные архитектуры предполагают распределение вычислительных функций между множеством узлов сети. Такой подход позволяет повысить устойчивость системы к отказам и уменьшить зависимость от центральных серверов. При этом возникает необходимость более сложных механизмов координации

и синхронизации между устройствами, что увеличивает требования к алгоритмам управления системой.

Гибридные архитектуры, основанные на интеграции облачных, туманных и периферийных вычислений, представляют собой наиболее перспективное направление развития IoT-систем. В рамках модели Cloud-Fog-Edge вычислительные задачи распределяются между различными уровнями инфраструктуры в зависимости от требований к скорости обработки данных и доступности вычислительных ресурсов. Такой подход позволяет снизить сетевые задержки, повысить эффективность использования ресурсов и обеспечить более гибкое масштабирование системы.

Использование многоуровневых вычислительных инфраструктур позволяет оптимизировать распределение нагрузки между облачными платформами, туманными узлами и периферийными устройствами, обеспечивая устойчивую работу системы в условиях роста объёмов данных и числа подключённых устройств.

Многоуровневая архитектура систем Интернета вещей

Архитектура систем Интернета вещей представляет собой многоуровневую структуру, объединяющую физические устройства, сетевую инфраструктуру и вычислительные платформы для обработки данных. Такая организация позволяет эффективно управлять большими потоками информации, поступающими от распределённых устройств, а также обеспечивать масштабируемость и устойчивость цифровых сервисов.

Современные IoT-системы, как правило, строятся на основе нескольких функциональных уровней, указанных на рисунке 2, каждый из которых выполняет определённые задачи по сбору, передаче, обработке и анализу данных. В большинстве исследований выделяются четыре ключевых уровня архитектуры: сенсорный, сетевой, уровень обработки данных и прикладной уровень [6].

Сенсорный уровень является базовым компонентом архитектуры Интернета вещей. Он включает различные устройства сбора данных, такие как датчики, исполнительные механизмы (актуаторы), RFID-метки и встроенные микроконтроллеры. Основной задачей данного уровня является взаимодействие с физической средой и преобразование измеряемых параметров в цифровые данные [7].

Современные сенсорные устройства часто оснащаются энергоэффективными микроконтроллерами и модулями беспроводной связи, что позволяет им работать автономно в течение длительного времени. Кроме того, активно развиваются технологии энергонезависимых

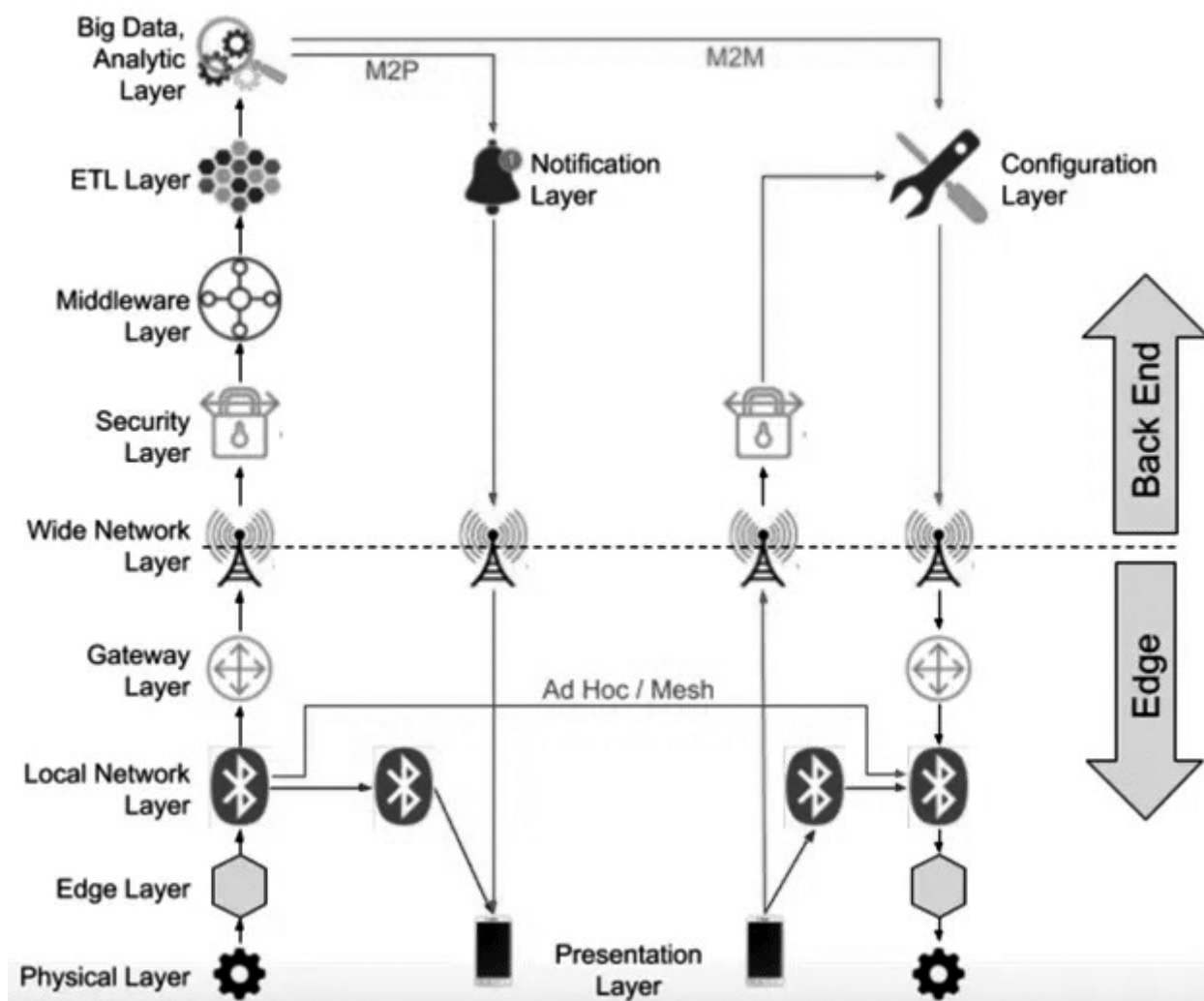


Рис. 2. Ключевые уровни архитектуры IoT

систем питания, основанные на использовании альтернативных источников энергии, таких как солнечная энергия или вибрационные генераторы [8].

Сетевой уровень обеспечивает передачу данных между устройствами Интернета вещей и вычислительными платформами. Он включает различные технологии связи, сетевые протоколы и коммуникационные инфраструктуры, обеспечивающие взаимодействие между элементами системы.

Для передачи данных в IoT широко применяются беспроводные технологии связи, такие как Wi-Fi, Bluetooth, Zigbee, LoRaWAN и NB-IoT. Каждая из этих технологий обладает своими характеристиками и используется в зависимости от требований к скорости передачи данных, дальности связи и энергопотреблению устройств [9].

Кроме того, в современных IoT-системах активно внедряются новые стандарты связи, включая сети пятого поколения (5G), обеспечивающие высокую пропускную способность и минимальные задержки передачи данных.

Уровень обработки данных отвечает за анализ, хранение и обработку информации, поступающей от устройств Интернета вещей. В современных системах используется распределённая модель вычислений, включающая периферийные (edge), туманные (fog) и облачные (cloud) ресурсы [10].

Edge-вычисления позволяют выполнять первичную обработку данных непосредственно на устройствах или в непосредственной близости от источников их генерации. Это снижает задержки передачи информации и уменьшает нагрузку на сетевую инфраструктуру.

Fog-уровень представляет собой промежуточную вычислительную инфраструктуру, расположенную между периферийными устройствами и облачными платформами. Его основная задача заключается в агрегации и предварительной обработке данных, поступающих от множества IoT-устройств.

Облачный уровень обеспечивает выполнение наиболее ресурсоёмких вычислений, включая анализ боль-

ших данных, обучение моделей машинного обучения и долгосрочное хранение информации. Использование облачных платформ позволяет масштабировать вычислительные ресурсы и обеспечивать обработку больших объёмов данных.

Прикладной уровень является верхним уровнем архитектуры IoT и отвечает за взаимодействие пользователей с системой. Он включает различные программные сервисы, интерфейсы и приложения, позволяющие анализировать данные и принимать управленческие решения.

На данном уровне реализуются различные цифровые сервисы, такие как системы умного города, интеллектуальные транспортные системы, промышленные системы автоматизации, телемедицинские платформы и системы управления энергопотреблением [11].

Прикладной уровень обеспечивает визуализацию данных, формирование аналитических отчётов и автоматизацию управленческих процессов. Использование методов искусственного интеллекта и анализа данных позволяет повысить эффективность работы таких систем и оптимизировать использование ресурсов.

Современные технологии и тенденции развития Интернета вещей

Активное развитие Интернета вещей сопровождается внедрением новых технологий обработки и передачи данных, а также интеграцией интеллектуальных алгоритмов управления. Современные исследования в данной области направлены на повышение автономности устройств, снижение задержек обработки информации и обеспечение безопасности распределённых систем.

Одним из ключевых направлений развития IoT является использование распределённых вычислительных архитектур, объединяющих облачные, туманные и периферийные вычисления. Такие архитектуры позволяют оптимально распределять вычислительную нагрузку между различными уровнями системы, обеспечивая более эффективную обработку данных и снижение сетевых задержек.

Современные IoT-платформы активно используют алгоритмы машинного обучения для решения задач прогнозирования, оптимизации энергопотребления, выявления аномалий и автоматизации процессов управления. Интеллектуальные системы способны адаптироваться к изменяющимся условиям среды и повышать эффективность функционирования цифровых сервисов.

Edge computing представляет собой подход, при котором обработка данных осуществляется непосредственно на периферийных устройствах или в непосред-

ственной близости от источников информации. Такой подход позволяет существенно снизить задержки передачи данных и повысить скорость реакции системы.

Fog computing выступает промежуточным уровнем между периферийными устройствами и облачными центрами обработки данных. Туманные узлы выполняют функции предварительной обработки, фильтрации и агрегации информации, поступающей от сенсоров. Это позволяет уменьшить объём данных, передаваемых в облако, и повысить эффективность использования сетевых ресурсов.

Совместное использование cloud-, fog- и edge-вычислений формирует распределённую вычислительную инфраструктуру, способную эффективно обрабатывать большие объёмы информации в условиях высокой динамики данных.

Использование IoT-датчиков совместно с цифровыми моделями обеспечивает возможность анализа состояния системы в реальном времени и повышает эффективность управления сложными техническими объектами.

Тем не менее применение блокчейна в IoT связано с рядом технических ограничений, включая высокие вычислительные требования и энергопотребление. В связи с этим ведётся активная разработка облегчённых алгоритмов консенсуса, предназначенных для работы в среде устройств с ограниченными ресурсами.

Эффективное функционирование IoT-экосистем требует использования единых стандартов и протоколов обмена данными. Среди наиболее распространённых протоколов можно выделить MQTT, CoAP и AMQP, которые обеспечивают надёжную передачу сообщений между устройствами и серверами.

Стандартизация протоколов взаимодействия играет важную роль в обеспечении совместимости различных устройств и платформ. В последние годы активно развивается стандарт Matter, предназначенный для обеспечения интероперабельности устройств разных производителей в системах умного дома.

Развитие стандартов и протоколов взаимодействия способствует формированию единой экосистемы Интернета вещей, позволяющей интегрировать устройства различных производителей в общую информационную инфраструктуру.

Заключение

Развитие Интернета вещей является одним из ключевых направлений цифровой трансформации современной экономики и технологической инфраструктуры.

Рост количества подключённых устройств, увеличение объёмов генерируемых данных и расширение сфер применения IoT требуют разработки эффективных архитектурных решений, обеспечивающих масштабируемость, надёжность и безопасность систем.

Проведённый анализ показал, что традиционные централизованные архитектуры не всегда способны обеспечить необходимую производительность при об-

работке больших потоков данных в реальном времени. В связи с этим всё большее распространение получают распределённые модели вычислений, основанные на интеграции облачных, туманных и периферийных вычислительных ресурсов. Архитектура Cloud-Fog-Edge позволяет оптимально распределять вычислительную нагрузку между различными уровнями системы, снижая задержки передачи данных и повышая эффективность обработки информации.

ЛИТЕРАТУРА

1. Kalan, R. A Survey on QoE Management Schemes for HTTP Adaptive Video Streaming: Challenges, Solutions, and Opportunities / R. Kalan, I. Dulger // IEEE Access. 2024. Vol.12. p. 170803–170839. <https://doi.org/10.1109/ACCESS.2024.3491613>
2. Mazhar, T. Quality of Service (QoS) Performance Analysis in a Traffic Engineering Model for Next-Generation Wireless Sensor Networks / T. Mazhar, M. Malik, S. Mohsan, Y. Li, I. Haq, S. Ghorashi, F. Karim, S. Mostafa // Symmetry. 2023. Vol. 15. Iss. 2. <https://doi.org/10.3390/sym15020513>
3. Kalyashina, A. Enhancing IoT systems through Cloud-Fog-Edge architectures challenges and opportunities / A. Kalyashina, Y. Smirnov, V. Esov, M. Kuznetsov, O. Dmitrieva // E3S Web of Conferences. 2024. Vol. 583. <https://doi.org/10.1051/e3sconf/202458306012>
4. Saovapakhiran, B. QoE-Driven IoT Architecture: A Comprehensive Review on System and Resource Management / B. Saovapakhiran, W. Naruephiphat, C. Charnsripinyo, S. Baydere, S. Özdemir // IEEE Access. 2022. Vol.10. p. 84579–84621. <https://doi.org/10.1109/ACCESS.2022.3197585>
5. Черепенин, В.А. Интеграция и оптимизация систем облачных, туманных и граничных вычислений: моделирование, задержки и алгоритмы / В.А. Черепенин, С.П. Воробьев // Известия высших учебных заведений. Северо-Кавказский регион. Технические науки. — 2024. — № 3(223). — С. 19–25. — DOI 10.17213/1560-3644-2024-3-19-25. — EDN LKVVKM
6. Das, R. A review on fog computing: Issues, characteristics, challenges, and potential applications / R. Das, M. M. Inuwa // Telematics and Informatics Reports. 2023. Vol. 10. <https://doi.org/10.1016/j.teler.2023.100049>
7. Tao W., Zhao L., Wang G., Liang R. Review of the internet of things communication technologies in smart agriculture and challenges. Computers and Electronics in Agriculture. 2021. No 189. Pp. 106–352.
8. Panahi, P.H.S. An Efficient Network-Based QoE Assessment Framework for Multimedia Networks Using a Machine Learning Approach / P.H.S. Panahi, A.H. Jalilvand, A. Diyanat // IEEE Open Journal of the Communications Society. 2025. Vol. 6. p. 1653–1669. <https://doi.org/10.1109/OJCOMS.2025.3543750>
9. Bourechak, A. At the Confluence of Artificial Intelligence and Edge Computing in IoT-Based Applications: A Review and New Perspectives / A. Bourechak, O. Zedadra, M. N. Kouahla, A. Guerrieri, H. Seridi, G. Fortino // Sensors. 2023. Vol. 23. Iss. 3. <https://doi.org/10.3390/s23031639>
10. Liu, B. A Survey of State-of-the-art on Edge Computing: Theoretical Models, Technologies, Directions, and Development Paths / B. Liu, Z. Luo, H. Chen, C. Li. // IEEE Access. 2022. Vol. 10. p. 54038–54063. <https://doi.org/10.1109/ACCESS.2022.3176106>
11. Li, Z. A quick response data collection and management system for digital twin production line based on cloud-fog-edge computing collaboration / Z. Li, X. Mei, D. Zhang, Z. Sun, J. Xu // Digital Twin. 2025. Vol.4. Iss.7. <https://doi.org/10.12688/digitaltwin.17907.2>

© Черепенин Валентин Анатольевич (blankzer@yandex.ru); Галушкина Дарья Владимировна (dashagaluschkina@yandex.ru);

Воробьев Сергей Петрович (vsp19999@yandex.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

МОДЕЛИРОВАНИЕ ДИНАМИКИ ПОТРЕБИТЕЛЬСКОГО СПРОСА НА ОСНОВЕ ИНТЕГРАЦИИ ЭНДОГЕННЫХ И ЭКЗОГЕННЫХ ФАКТОРОВ С УЧЕТОМ ИХ ЛАГОВОГО ВОЗДЕЙСТВИЯ

MODELLING THE DYNAMICS OF CONSUMER DEMAND BASED ON THE INTEGRATION OF ENDOGENOUS AND EXOGENOUS FACTORS, TAKING INTO ACCOUNT THEIR LAG EFFECTS

M. Shibichenko

Summary. This study examines issues related to forecasting the dynamics of consumer demand under the influence of external and internal determinants with delayed effects. It briefly describes the dynamics of factors influencing consumer demand in Russia over the past few years. The article proposes an original forecasting algorithm that integrates exogenous and endogenous demand factors into a single dynamic structure. From a mathematical perspective, the approach to separating the instantaneous and cumulative effects of prices and incomes is formalised, which will improve the accuracy of forecasting.

Keywords: demand, forecasting, model, accuracy, time lag, delayed effect.

Шибиченко Михаил Иванович

аспирант, Московский финансово-юридический
университет МФЮА
shibmish@yandex.ru

Аннотация. Исследование посвящено изучению вопросов, связанных с прогнозированием динамики потребительского спроса в условиях действия внешних и внутренних детерминант с отсроченным действием. Дана краткая оценка изменений факторов во времени, влияющих на потребительский спрос в России за последние несколько лет. В статье предложен авторский алгоритм прогнозирования, который объединяет экзогенные и эндогенные факторы спроса в единую динамическую структуру. С математической точки зрения формализован подход разделения мгновенных и накопленных эффектов влияния цен и доходов, что позволит повысить точность прогнозирования.

Ключевые слова: спрос, прогнозирование, модель, точность, временной лаг, отложенное действие.

Прогноз спроса — ключевой элемент управленческих решений. Он задаёт исходные ориентиры как для текущих действий, так и для долгосрочного планирования. От точности таких оценок зависят многие экономические исследования. В промышленной организации это позволяет оценивать последствия слияний и запуск новых продуктов, анализировать стратегию фирм и их реакцию на конкуренцию. В международной торговле расчёт спроса помогает измерять эффект тарифов и выявлять выгоды от внешнеэкономических операций и внутренней миграции [1].

Современная структура спроса усложняется. На него одновременно влияют внутренние и внешние факторы, и их взаимодействие далеко не линейно. К эндогенным драйверам относятся доходы домохозяйств, инфляционные ожидания и общие потребительские настроения. Экзогенные импульсы формируют макроэкономические шоки и крупные непредвиденные события: глобальные финансовые кризисы, пандемия COVID-19, а также санкционные ограничения [2].

В результате экономическая среда становится менее стабильной. И это заметно усиливает изменчивость поведения потребителей. Их реакция на изменения в эко-

номике часто проявляется не сразу, а с определённой задержкой. Такие лаговые эффекты усложняют анализ и прогнозирование спроса [3]. И в данном случае самым большим препятствием является масштаб прогнозирования сотен или тысяч различных временных рядов с разнообразными особенностями.

В таблице 1 представлена сравнительная характеристика ключевых показателей, отражающих изменения в динамике и структуре потребительского спроса в России за период 2010–2025 гг.

Анализ данных из таблицы 1 показывает, что потребительский спрос в России стал гораздо более нестабильным. Именно поэтому нужны новые методы, которые смогут одновременно учитывать и внутренние, и внешние факторы, и то, что их эффект проявляется с опозданием. Высокая актуальность и практическая значимость отмеченных вопросов обусловили выбор темы данной статьи.

В настоящее время наиболее распространенными подходами к прогнозированию временных рядов являются модель экспоненциального сглаживания, модель Холта-Уинтера, модель Бокса и Дженкинса, регрессион-

Таблица 1.
Динамика факторов, влияющих на потребительский
спрос в России¹

Показатель / аспект	2010–2015	2016–2020	2021–2025	Тенденция / вывод
Волатильность потребитель- ского спроса (индекс, %)	6–9 %	10–15 %	15–25 %	Сущест- вен- ный рост нестабиль- ности спроса
Влияние экзо- генных факто- ров (кризисы, пандемия, санкции)	Умеренное (кризис 2014–2015)	Высокое (COVID-19)	Очень высокое (санкции, инфляция, геополитика)	Усиление внешнего воздействия
Доля онлайн- потребления (%)	2–5 %	5–12 %	15–25 %	Быстрая цифровиза- ция спроса
Средняя дли- тельность лагов реакции спроса (месяцы)	1–2	2–4	3–6	Увеличение временных задержек
Точность классических моделей (ошиб- ка прогноза, %)	10–15 %	15–25 %	25–35 %	Снижение эффектив- ности тра- диционных моделей
Необходимость учета эндоген- ных факторов (доходы, ожидания)	Средняя	Высокая	Критическая	Усиление внутренней динамики
Использование комплексных моделей (с лагами и факторами), % компаний	10–20 %	20–35 %	40–65 %	Рост вне- дрения про- двинутой аналитики

ная модель и модели ARIMA. Особенности их использо-
вания детально раскрыты в трудах Габдулина Р.Р., Лясков-
ской Е.А., Коровина А.М., Реца Е.А., Fang Li, Tao Li, Oliver
Nordemann, Rainer Schlosser.

¹ Источник: составлено автором по данным Федеральная
служба государственной статистики «Социально-экономическое
положение России» (2010–2024 гг.), Банк России «Доклад о денеж-
но-кредитной политике» (2015–2025 гг.), «Основные направления
единой государственной денежно-кредитной политики» (разные
годы), Ассоциация компаний интернет-торговли «Развитие интер-
нет-торговли в России» (2015–2024 гг.)

Методы многомерного прогнозирования спроса с по-
мощью технологий искусственного интеллекта описы-
вают Степанищев С.А., Турищев Д.К., Петрич А.С., Anitha S.,
Neelakandan R., Bui Bich Xuan, Erlend Dancke Sandorf.

Несмотря на активный интерес ученых к рассматри-
ваемой проблематике не все вопросы в данном пред-
метном поле еще решены. Например, недостаточно
изучены механизмы совместного влияния внутренних
и внешних факторов на поведение потребителей в усло-
виях нестабильной экономической среды. Кроме того,
в дальнейшем развитии нуждается методика учета вре-
менных лагов, которые искажают оценку силы и направ-
ленности этих воздействий.

Таким образом, цель статьи заключается в рассмо-
трении особенностей моделирования динамики потре-
бительского спроса на основе интеграции эндогенных
и экзогенных факторов с учетом их лагового воздей-
ствия.

Первым шагом в процессе моделирования динамики
потребительского спроса является проведение иденти-
фикации структуры факторов и предварительной обра-
ботки временных рядов.

Для этого необходимо сформировать репрезен-
тативную выборку предикторов. Главное требование
к ним — способность адекватно описать волатильность
потребительского спроса. На данном этапе проводится
не просто группировка данных, а проверка их на соот-
ветствие условиям применимости динамического мо-
делирования. Будем исходить из постулата, что текущий
спрос Q_t является функцией не только текущих стимулов,
но и накопленного опыта потребителя, зафиксирован-
ного в лаговых значениях.

В рамках авторского подхода выделены две ключе-
вые группы факторов.

1. Эндогенные факторы (управляемые внутри си-
стемы), которые характеризуются высокой частотой
обновления и непосредственным влиянием
на объем реализации продукции.
2. Экзогенные факторы (внешние по отношению
к моделируемой системе) определяют общую ем-
кость рынка и его макроэкономический контекст.

Особенностью разрабатываемого алгоритма модели-
рования спроса является использование дифференци-
рованных лагов для каждой группы. Для цены лаг будет
минимальным ($t-1$). В то время как для макроэкономи-
ческих показателей (например, располагаемых доходов)
он может составлять до нескольких кварталов ($t-k$).

Прежде чем включать лаговые значения в модель,
необходимо убедиться в стационарности рядов, чтобы

избежать феномена «ложной корреляции». Для первичного анализа временного ряда важно проверить стационарность. Чаще всего используют расширенный тест Дики-Фуллера (ADF) [4]. Если обнаружен единичный корень, ряд нестационарен. В этом случае применяют дифференцирование и переходят к первым разностям:

$$\Delta X_t = X_t - X_{t-1}$$

Далее следует определить лаговую структуру. Здесь помогают автокорреляционная и частная автокорреляционная функции. Они указывают, какие задержки существенны. Параметры p и q выбирают так, чтобы учесть память процесса и не перегрузить модель. Итогом становится спецификация, которая сочетает достаточную гибкость и экономность по числу параметров. Технические ориентиры для этой проверки сведены в обобщенный перечень, представленный в таблице 2.

Таблица 2.

Комплект диагностических процедур для начальной проверки и подготовки данных²

Процедура этапа	Математический инструментарий	Формальный критерий принятия решения	Ожидаемый результат для модели
Проверка на стационарность	Расширенный тест Дики-Фуллера	$t_{stat} < t_{crit}$ (отклонение H0)	Подтверждение отсутствия единичного корня
Определение глубины лага	Информационный критерий Шварца	min(BIC) по сетке лагов	Оптимальное число периодов запаздывания
Анализ сонаправленности	Тест на коинтеграцию Йохансена	Наличие коинтеграционных векторов ($r > 0$)	Возможность построения долгосрочной модели
Фильтрация шума	Сезонная декомпозиция	Коэффициент детерминации тренда $R^2 > 0,85$	Выделение чистой циклической компоненты

Затем для каждого выбранного предиктора X вычисляется кросс-корреляция с целевой переменной спроса Q . Эта процедура фиксирует, как спрос откликается на разные стимулы. Картина получается предметной: видны и направление, и задержка реакции. На примере цен это особенно наглядно. Изменение цены не всегда действует мгновенно. Эффект часто распределён во времени. Его удобно описывать вектором лагов, который получают из профиля кросс-корреляции

$$CCF(k) = \frac{\sum_{t=1}^n (X_{t-k} - \bar{X})(Q_t - \bar{Q})}{\sqrt{\sum_{t=1}^n (X_t - \bar{X})^2 \sum_{t=1}^n (Q_t - \bar{Q})^2}}$$

где k — величина временного сдвига.

² Составлено автором

Данный расчет позволяет отсеять те периоды, где корреляция попадает в доверительный интервал (белый шум), тем самым оптимизируя структуру будущего уравнения регрессии и предотвращая переобучение модели.

Финальным действием первого этапа является построение матрицы, в которой для каждого фактора определена сила его влияния в системе потребительского спроса. На основе автокорреляционной и частной автокорреляционной функции выстраивается структура модели авторегрессионного распределенного лага.

В таблице 3 приведен пример матрицы для синтеза эндогенных и экзогенных параметров.

Таблица 3.

Матрица лаговых структур переменных (пример первичного формирования)³

Переменная	Оптимальный лаг (k)	Оценка влияния	Обоснование включения
Цена (P)	$k = 0,1$	0,89	Немедленная реакция + эффект памяти
Доход (I)	$k = 3$	0,45	Квартальный цикл пересмотра расходов
Спрос (Q)	$k = 1,12$	0,72	Инерция + годовая сезонность

Второй шаг моделирования динамики потребительского спроса предполагает построение динамической модели.

После того как были определены значимые временные сдвиги, наступает этап непосредственного конструирования расчетной модели. В научном поиске важно избежать излишнего упрощения, поэтому будем использовать структуру авторегрессионного распределенного лага, о которой уже отмечалось ранее. Это позволяет одновременно учитывать и внешние шоки, и внутреннюю инерцию самого покупательского поведения [5]. Таким образом логика разрабатываемой модели заключается в следующем: текущий объем реализации выступает производной от накопленного рыночного опыта и структуры спроса в предшествующих периодах.

Итак, объединим текущие значения цен и доходов с их прошлыми состояниями. Каждому фактору присвоим свой весовой коэффициент, который показывает силу его влияния на итоговый спрос. Общий вид модели, адаптированной под задачу интеграции факторов, представлен формулой:

$$Q_t = \alpha + \sum_{i=0}^p \beta_i P_{t-i} + \sum_{j=0}^q \gamma_j I_{t-j} + \lambda Q_{t-1} + \epsilon_t$$

³ Составлено автором

где Q_t — объем спроса в текущий момент; P и I — цена и доход соответственно; β , γ , λ — параметры, которые нужно вычислить, ϵ_t — стохастическая составляющая (случайная ошибка), отражающая влияние неучтенных в модели факторов/

В таблице 4 описано какую роль играет каждый блок этой формулы в логике алгоритма.

Таблица 4.

Функциональные блоки динамической модели спроса⁴

Компонент формулы	Техническое название	Эконометрическая интерпретация компонента	Роль в прогнозе
α	Свободный член	Базовый уровень спроса, который есть всегда, даже если факторы на нуле	Точка отсчета
$\beta_i P_{t-1}$	Блок ценового лага	Показывает, как сильно шоки старых цен влияют на выбор покупателя	Оценка чувствительности
$\gamma_j I_{t-j}$	Блок внешних факторов	Отражает влияние зарплат и экономики, которое доходит до рынка с задержкой	Учет макросреды
λQ_{t-1}	Авторегрессия	Учитывает привычку: если товар популярен, он продает сам себя за счет инерции	Оценка стабильности

После того, как структура модели определена, алгоритм переходит к нахождению конкретных чисел. Дальнейшая оценка параметров проводится методом наименьших квадратов. Однако требуется одна принципиальная поправка: лаговые эффекты должны быть разделены так, чтобы их вклад не дублировался и отражал фактическую динамику системы. Для этого анализируют эластичности [6]. Анализ позволяет показать на сколько процентов изменится спрос при росте цены или дохода, и как быстро проявится этот эффект.

Краткосрочные и долгосрочные влияния различаются. Вначале оцениваются локальные реакции по отдельным лагам, затем их последовательно агрегируют, чтобы получить суммарные эффекты. Такая поэтапная свертка коэффициентов эластичности обеспечивает сопоставимость результатов и устойчивость интерпретаций. Более детально данная процедура изложена в таблице 5.

Отдельно требуется учесть внешние шоки, которые нарушают обычный ход динамики. Резкие изменения курса, праздничные периоды, административные решения — все это примеры событий, не подчиняющихся регулярной структуре. Их вводят в спецификацию в виде фиктивных переменных. Такие индикаторы помогают

Таблица 5.

Процедура агрегирования лаговых эластичностей для оценки краткосрочных и долгосрочных эффектов факторов на спрос⁵

Тип эффекта	Расчетный алгоритм (формула)	Эконометрическое содержание	Прогнозная интерпретация
Мгновенный	Коэффициент β_0	Отражает первичную реакцию спроса в момент воздействия фактора	Оценка эффективности оперативных маркетинговых мер
Накопленный	$\sum_{i=0}^p \beta_i$	Суммарный объем отклика системы за весь период лагового воздействия	Анализ «инерционного хвоста» после завершения активной фазы
Долгосрочный	$\frac{\sum \beta_i}{1 - \lambda}$	Уровень стабилизации спроса при переходе системы в новое равновесие	Стратегический прогноз рыночной позиции в долгосрочном периоде

модели сохранять стабильность и корректно описывать периоды аномальных отклонений [7].

Финальный шаг — валидация. Сначала необходимо проанализировать ϵ_t . Важный критерий — отсутствие систематических зависимостей. Для проверки используют тесты Бреуша–Годфри и Дарбина–Уотсона [8]. Если автокорреляции нет, временная структура учтена корректно, и модель извлекает основную информацию из данных. Затем проверяем мультиколлинеарность по VIF (Variance Inflation Factor). Это принципиально: цена и доход часто движутся вместе. При высокой взаимосвязи оценки коэффициентов искажаются, экономическая интерпретация теряется.

Точность прогноза оценивается по MAPE. Для этого сопоставляются расчётные и фактические значения. Допустимое значение ошибки около 10–15 %. При таком уровне прогноз пригоден для практики, включая планирование.

Наблюдаемая динамика спроса характеризуется высокой изменчивостью. Макроэкономические возмущения повышают дисперсию и изменяют структуру автокорреляций, что делает статистические спецификации методологически несостоятельными для описания процессов с запаздывающей реакцией.

В работе сформулирован и реализован алгоритм моделирования, объединяющий эндогенные и экзогенные детерминанты в рамках авторегрессионной струк-

⁴ Составлено автором

⁵ Составлено автором

туры с распределёнными лагами. Такая постановка позволяет: (i) идентифицировать краткосрочные отклики и долгосрочные накопленные эффекты изменений цен и доходов; (ii) обеспечить экономическую интерпретируемость параметров за счёт явного разнесения временных контуров влияний; (iii) повысить робастность оценок к структурным сдвигам за счёт встроенной лаговой динамики.

Эмпирические результаты указывают на улучшение прогностических характеристик и устойчивости спец-

ификаций при контроле задержанных реакций. Практически это выражается в снижении ошибки прогнозирования и повышении согласованности оценок эффектов по горизонтам. Существенным преимуществом предлагаемой модели является учёт инерционности поведения потребителей и способность адаптироваться к изменению режимов, что повышает адекватность описания рыночной динамики в условиях внешних шоков и внутренней негладкости процессов.

ЛИТЕРАТУРА

1. Ткаченко В.В., Волков Е.О., Кужильный А.В. Информационная система оценки потребительского спроса и прогнозирования результатов рекламных кампаний и промоакций // Естественнo-гуманитарные исследования. 2024. № 4 (54). С. 247–253.
2. Донскова Д.А., Погребняк А.Я., Ворокова Н.Х. Прогнозирование спроса на продукцию «Золотое яблоко» с использованием адаптивных моделей временных рядов // Финансовая экономика. 2025. № 5. С. 42–45.
3. Oliver Nordemann, Rainer Schlosser Long-term consumer selection strategies for a heterogeneous consumer base when demand depends on purchase and advertising histories // International Transactions in Operational Research. 2025. Volume 33, Issue 1. P. 76–83.
4. Xin Yi Construction of Online Agricultural Product Consumption Demand Analysis Model Based on Data Mining Technology // Advances in Multimedia. 2022. Volume 20, Issue 19. P. 65–73.
5. Ратнер С.В., Шапошников А.М. Прогнозирование изменений потребительского поведения в условиях экономического кризиса // Экономический анализ: теория и практика. 2022. Т. 21. № 5 (524). С. 911–926.
6. Anitha S., Neelakandan R. Demand Forecasting New Fashion Products: A Review Paper // Journal of Forecasting. 2024. Volume 44, Issue 2. P. 111–118.
7. Цебрено К.Н., Ермолич Д.А. Метод Хольта-Винтерса в системе оперативного прогнозирования в крупном продуктовом ритейле // Вестник ИМСИТ. 2025. № 4 (104). С. 41–46.
8. Harsha Chamara Hewage, H. Niles Perera, Kasun Bandara Enhancing Demand Forecasting in Retail: A Comprehensive Analysis of Sales Promotional Effects on the Entire Demand Life Cycle // Journal of Forecasting. 2025. Volume 45, Issue 1. P. 56–62.

© Шибиченко Михаил Иванович (shibmish@yandex.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

ДИАГНОСТИЧЕСКАЯ ИНФОРМАТИВНОСТЬ БИОМАРКЕРОВ ПРИ ОЦЕНКЕ ВОСПАЛИТЕЛЬНЫХ И РЕПАРАТИВНЫХ ПРОЦЕССОВ ПОЛОСТИ РТА

DIAGNOSTIC INFORMATIVENESS OF BIOMARKERS IN ASSESSING INFLAMMATORY AND REPARATIVE PROCESSES IN THE ORAL CAVITY

**N. Erokina
A. Mironov
N. Zakharova
A. Parfenov
P. Dzhagaryan
G. Bakhteeva
S. Savelyeva
D. Balakina**

Summary. The purpose of this study was to establish the diagnostic significance of cytokines/chemokines and growth factors in assessing the reparative processes of the tooth socket in patients with type 2 diabetes mellitus. **Materials and methods of research.** 50 patients who underwent tooth extraction due to chronic periodontitis were examined. 25 of them had type 2 diabetes. The patients' clinical parameters, as well as the level of cytokines/chemokines and growth factors were studied by enzyme immunoassay. **The results and their discussion.** In patients with type 2 diabetes mellitus, there was an imbalance between pro- and anti-inflammatory cytokines/chemokines and a decrease in the content of growth factors in the recess of the extracted tooth. However, clinical data indicated a slowdown in the reparative processes in the tooth socket. **Conclusion.** According to the dynamics of biomarkers of the tooth socket, an inadequate response of the immune system was established, which caused a violation of the repair processes in patients with type 2 diabetes.

Keywords: biomarkers, cytokines, growth factors, tooth socket, diabetes.

Ерокина Надежда Леонидовна

д.м.н., профессор, ФГБОУ ВО Саратовский
государственный медицинский университет
им. В.И. Разумовского Минздрава России
nadleo@mail.ru

Миронов Алексей Юрьевич

к.м.н., ФГБОУ ВО Саратовский государственный
медицинский университет им. В.И. Разумовского
Минздрава России

Захарова Наталья Борисовна

д.м.н., профессор, ФГБОУ ВО Саратовский
государственный медицинский университет
им. В.И. Разумовского Минздрава России

Парфенов Артем Кириллович

ассистент, ФГБОУ ВО Саратовский государственный
медицинский университет им. В.И. Разумовского
Минздрава России

Джагарян Папик Джагарович

аспирант, ФГБОУ ВО Саратовский государственный
медицинский университет им. В.И. Разумовского
Минздрава России

Бахтеева Галия Рифатовна

к.м.н., доцент, ФГБОУ ВО Саратовский государственный
медицинский университет им. В.И. Разумовского
Минздрава России

Савельева Светлана Станиславовна

к.м.н., доцент, ФГБОУ ВО Саратовский государственный
медицинский университет им. В.И. Разумовского
Минздрава России

Балакина Диана Дмитриевна

ассистент, ФГБОУ ВО Саратовский государственный
медицинский университет им. В.И. Разумовского
Минздрава России

Аннотация. Целью настоящего исследования явилась установка диагностической значимости цитокинов/хемокинов и факторов роста в оценке репаративных процессов лунки зуба у больных сахарным диабетом 2 типа. **Материалы и методы исследования.** Обследовано 50 больных, которым проводилось удаление зуба по причине хронического периодонтита. У 25 из них был сахарный диабет 2 типа. У пациентов изучали клинические показатели, а также уровень цитокинов/хемокинов и факторов роста методом иммуноферментного анализа. **Результаты и их обсуждение.** У пациентов сахарным диабетом 2 типа отмечался дисбаланс между про- и противовоспалительными цитокинами/хемокинами и снижение содержания факторов роста в лунке удаленного зуба. При этом клинические данные указывали на замедление репаративных процессов в лунке зуба. **Заключение.** По динамике биомаркеров лунки зуба установлен неадекватный ответ иммунной системы, явившийся причиной нарушения процессов репарации у больных СД 2 типа.

Ключевые слова: биомаркеры, цитокины, факторы роста лунка зуба, сахарный диабет.

Биомаркеры широко применяются в различных областях медицины, в том числе стоматологии. Они используются с диагностической целью и для оценки течения и эффективности лечения при различных заболеваниях. Влияние сахарного диабета (СД) на заболевания полости рта, развитие и нарушение течения воспалительных и регенераторных процессов подтверждено многими исследованиями [3, 6, 7, 8]. Одной из причин развития воспалительных заболеваний, приводящих к замедлению процессов регенерации, является дисбаланс между про- и противовоспалительными пулами цитокинов [1, 2, 4, 5].

Цель исследования — установить влияние цитокинов/хемокинов и факторов роста на течение репаративных процессов в лунке зуба у больных сахарным диабетом 2 типа.

Материал и методы исследования

На базе кафедры стоматологии хирургической и челюстно-лицевой хирургии ФГБОУ ВО «Саратовский ГМУ им. В.И. Разумовского» Минздрава России выполнено ретроспективное исследование. Обследованы 50 пациентов (23 мужчины и 27 женщин), из них у 25 пациентов был СД 2 типа. *Критерии включения* в исследование: средний возраст (46–60 лет), простое удаление зуба при диагнозе хронический периодонтит, СД 2 типа в стадии компенсации (целевой уровень гликированного гемоглобина <7,5). *Критерии исключения* декомпенсированная форма сопутствующих заболеваний, неудовлетворительная гигиена полости рта. Сравнение проводилось с группой здоровых людей (25 человек), у которых исследовалась десневая жидкость. В исследовании применялась блочная рандомизация. Пациентам с хроническим периодонтитом проводились оценка клинических показателей, характеризующих воспалительный и репаративный процессы — боли, гиперемии слизистой оболочки в месте удаления зуба, сроков эпителизации лунки. Методом иммуноферментного анализа определяли уровень цитокинов/хемокинов (IL-1 β , IL-6, IL-8, MCP1, RAIL-1) и факторов роста (VEGF, TGF β 1) в лунках на следующие и 5-е сутки после удаления зуба. Определение уровня цитокинов и фактора роста VEGF проводили методом твердофазного иммуноферментного анализа с использованием коммерческих наборов реагентов АО Вектор-Бест, Новосибирск. Для исследования TGF- β 1 применяли трех-стадийный «сэндвич»-вариант ИФА с использованием моно- и поликлональных антител (R&D Systems, Великобритания). Исследования проведены в соответствии со стандартами надлежащей клинической практики (GoodClinicalPractice) и принципами Хельсинской декларации. Статистическая обработка данных проведена стандартными способами, использовали медиану и интерквартильный интервал.

Результаты исследования

У пациентов без нарушения углеводного обмена содержание биомаркеров в лунках зубов на следующий день после удаления свидетельствует о развитии умеренного воспаления. На это указывает повышение уровня провоспалительных цитокинов/хемокинов: IL-1 β до 24,73 пг/мл, IL-6 до 157 пг/мл, IL-8 до 588,8 пг/мл, MCP1 до 33,8, относительно группы здоровых людей (IL-1 β — 5,9 пг/мл; IL-6 — 8,8 пг/мл; IL-8 — 95,7 пг/мл; MCP1 — 20,35 пг/мл), уменьшение противовоспалительного цитокина RAIL-1 до 1987,3 пг/мл (в группе сравнения — 3669,9 пг/мл). Также отмечаются отличия в содержании факторов роста: повышение уровня VEGF до 116,2 пг/мл и снижение TGF β 1 до 15,2 пг/мл (в группе сравнения VEGF — 29,3 пг/мл; TGF β 1 — 24,2 пг/мл). На 5-е сутки после удаления отмечается изменение цитокинового статуса, указывающее на уменьшение степени воспаления в лунке — снижается уровень провоспалительных цитокинов/хемокинов: IL-1 β до 15,4 пг/мл, IL-6 до 138,8 пг/мл, IL-8 до 521,8 пг/мл, MCP1 до 17,2 пг/мл. При этом наблюдается увеличение уровня противовоспалительного цитокина RAIL-1 до 2726,7 пг/мл. Так же отмечается некоторое повышение уровня факторов роста: VEGF до 127,9 пг/мл и TGF β 1 до 25,1 пг/мл.

У больных СД 2 типа, на следующие сутки после удаления зуба, уровень провоспалительных цитокинов/хемокинов был ниже, чем у пациентов без нарушения углеводного обмена и равен: IL-1 β — 10,32 пг/мл, IL-6 — 52,75 пг/мл, IL-8 — 531,17 пг/мл, MCP1 — 35,9 пг/мл. Уровень противовоспалительного цитокина RAIL-1 был равен 1595,34 пг/мл, что ниже, чем в группе больных без нарушений углеводного обмена. Уровень фактора роста VEGF был ниже (92,8 пг/мл), а уровень TGF β 1 (16,8 пг/мл) значимо не отличался от значений группы пациентов без нарушений углеводного обмена. На 5-е сутки после удаления зуба уровень провоспалительных цитокинов/хемокинов составил: IL-1 β — 6,02 пг/мл, IL-6 — 116,27 пг/мл, IL-8 — 426,67 пг/мл, что превышало аналогичные значения группы пациентов без нарушений углеводного обмена, за исключением MCP1, уровень которого (19,6 пг/мл) достоверно не отличался. Содержание противовоспалительного цитокина RAIL-1 к 5-м суткам увеличилось (2290,67), но было ниже, чем у пациентов без нарушений углеводного обмена. Уровень факторов роста так же повышался относительно 1 суток после удаления и был меньше, чем у пациентов без нарушения углеводного обмена: VEGF — 110,47 пг/мл, TGF β 1 — 20,2 пг/мл.

При этом динамика клинических показателей, характеризующих активность воспалительного процесса в лунке после удаления зуба отличалась в группах пациентов с СД 2 типа и не имеющих нарушений углеводного обмена. Так, на следующие сутки после удаления больные СД 2 типа, по данным вербальной ранговой, отсутствие боли отмечали в 2,7 раза реже и у трети пациентов

сахарным диабетом 2-го типа, отмечавших боль, она была умеренной или сильной, в то время как в группе сравнения — только слабая. На 5-ые сутки слабая или умеренная боль в области лунки зуба наблюдалась только в группе пациентов с СД 2 типа. При осмотре, на следующие после удаления зуба сутки у больных СД 2 типа в 2 раза чаще, относительно группы сравнения, отмечалась гиперемия десны вокруг лунки. При этом, в отличие от пациентов без нарушений углеводного обмена, в половине случаев она распространялась на маргинальную, прикрепленную десну или переходную складку и сохранялась (уменьшаясь) на 5 сутки. Сроки эпителизации так же отличались в описанных группах пациентов. У больных СД 2 типа краевая эпителизация лунки зуба на 5-ые сутки определялась 1,7 раз реже, чем в группе сравнения, а полная эпителизация происходила в более поздние сроки. Воспалительные осложнения в виде альвеолита отмечались только у 2 (8 %) пациентов с сахарным диабетом 2-го типа.

Обсуждение

Процессы заживления лунки зуба характеризуются развитием типичной реакции заживления ран, при которой происходит смена физиологических регенеративных процессов. Важную роль при этом играет взаимодействие мезенхимальных и эпителиальных клеток, которое регулируется веществами, называемыми биомаркерами, оказывающими местное и системное действие. К ним относятся цитокины и факторы роста. Раневой процесс включает несколько фаз [2, 3, 6, 8]. В фазу воспаления отторгаются омертвевшие клетки и уничтожаются микроорганизмы. Эта фаза имеет экссудативный и резорбтивный период. В экссудативный период формируется сгусток из тромбоцитов и эритро-

цитов, опутанный сетью белков плазмы. Этот комплекс клеток и фибрина служит резервуаром факторов роста, направляет рост сосудов и привлекает клетки в рану. В резорбтивный период происходит очищение раны и инициируется иммунная реакция. Восстановительная фаза имеет пролиферативный и репаративный периоды. Проллиферативный характеризуется вращением в сгусток капилляров, формированием грануляционной ткани и эпителизацией краев раны. Репаративный период — финальный период заживления раны с полной ее эпителизацией и реорганизацией рубца [2, 7, 8, 9]. В раневом процессе активное участие принимают вырабатываемые клетками цитокины, хемокины и факторы роста. Физиологические процессы нарушаются у пациентов с сахарным диабетом [1, 2, 6, 9]. Динамика содержания провоспалительных цитокинов и факторов роста свидетельствует о замедленном протекании воспалительной реакции, то есть неадекватном ответе иммунной системы у больных СД 2 типа. Выявленные нами нарушения удлиняют сроки течения репаративных процессов, а в ряде случаев приводят к развитию осложнений.

Заключение

Полученные данные свидетельствуют о возможности использования содержания изученных биомаркеров для диагностики замедленного протекания воспалительной реакции и установления неадекватного ответа иммунной системы, у пациентов с СД 2 типа. Он характеризуется дисбалансом между про- и противовоспалительными пулами цитокинов и снижением содержания факторов роста в лунке после удаления зуба. Это приводит к замедлению репаративных процессов, что подтверждают клинические данные (продолжительность боли, гиперемии и эпителизации).

ЛИТЕРАТУРА

1. Миронов А.Ю., Мартынова М.И., Ерокина Н.Л., Лепилин А.В. Клинико-лабораторное обоснование использования биомаркеров для диагностики воспалительных и репаративных процессов полости рта у больных сахарным диабетом // Волгоградский научно-медицинский журнал. 2024. Т. 21, №2. С.84–87.
2. Ерокина Н.Л., Лепилин А.В., Миронов А.Ю., Захарова Н.Б., Фищев С.Б. Роль цитокинов, хемокинов и факторов роста в заживлении лунок удаленных зубов // Пародонтология. 2021. Т. 26, №1. С. 58–63. DOI:10.33925/1683-3759-2021-26-1-58-63
3. Лепилин А.В., Ерокина Н.Л., Иванова И.А., Миронов А.Ю., Григорьева Д.А., Мартынова М.И. Диагностическое значение исследования десневой жидкости у пациентов с несъемными конструкциями зубных протезов // Современная ортопедическая стоматология. 2018. №29. С.13–16.
4. Мартынова М.И., Миронов А.Ю., Лепилин А.В., Ерокина Н.Л. Диагностическая информативность цитокинов/хемокинов и факторов роста в стоматологии // Волгоградский научно-медицинский журнал. 2024. Т.21, №2. С. 26–30.
5. Zhao R., Liang H., Clarke E., Jackson C., Xue M. Inflammation in chronic wounds. // Int J Mol Sci. 2016;17(12):2085. DOI:10.3390/ijms17122085
6. Park J.W., Hwang S.R., Yoon I.S. Advanced growth factor delivery systems in wound management and skin regeneration. // Molecules. 2017;22:1259. DOI:10.3390/molecules22081259
7. Olczyk P., Mencner Ł., Komosińska-Vashev K. The role of the extracellular matrix components in cutaneous wound healing. // Bio Med Res Int. 2014;747584:1–8. DOI:10.1155/2014/747584
8. Barrientos S., Brem H., Stojadinovic O., Tomic-Canic M. Clinical application of growth factors and cytokines in wound healing. // Wound Repair Regen. 2014;22(5):569–578. DOI:10.1111/wrr.12205
9. Gilbert R.W.D., Vickaryous M.K., Vitoria-Petit A.M. Signalling by transforming growth factor beta isoforms in wound healing and tissue regeneration. // J Dev Biol. 2016;4(2):21. DOI:10.3390/jdb4020021

© Ерокина Надежда Леонидовна (nadleo@mail.ru); Миронов Алексей Юрьевич; Захарова Наталья Борисовна; Парфенов Артем Кириллович; Джагарян Папик Джагарович; Бахтеева Галия Рифатовна; Савельева Светлана Станиславовна; Балакина Диана Дмитриевна
Журнал «Современная наука: актуальные проблемы теории и практики»

СОВРЕМЕННОЕ СОСТОЯНИЕ ПРОБЛЕМЫ ЛЕЧЕНИЯ ЭНДОМЕТРИОМ: ПРОТИВОРЕЧИЯ, РИСКИ И ПЕРСПЕКТИВЫ ПЕРСОНАЛИЗИРОВАННОЙ ТЕРАПИИ

THE CURRENT STATE OF THE ENDOMETRIOMA TREATMENT PROBLEM: CONTRADICTIONS, RISKS AND PROSPECTS FOR PERSONALIZED THERAPY

**V. Kotelnikov
L. Tselkovich
R. Balter
T. Ivanova
O. Ilchenko**

Summary. A comprehensive analysis of modern approaches to the diagnosis and treatment of endometriomas, as well as their impact on ovarian reserve and fertility, has been conducted. The key controversies that arise when choosing a treatment strategy have been reviewed. A systematic review of the risks of iatrogenic ovarian damage during cystectomy has been conducted, and the effectiveness and safety of alternative methods, including ethanol sclerotherapy and ablative techniques, have been reviewed.

Keywords: endometrioma, endometriosis, ovarian reserve, cystectomy, sclerotherapy, fertility, personalized therapy, anti-Müllerian hormone.

Котельников Виталий Владимирович

Соискатель,

Самарский государственный медицинский университет

Kotelnikov.v.v@ibioclinic.com

Целкович Людмила Савельевна

Д.м.н., профессор,

Самарский государственный медицинский университет

РФ, г. Самара

Балтер Регина Борисовна

Д.м.н., профессор,

Самарский государственный медицинский университет

samaraobsgyn2@yandex.ru

Иванова Татьяна Владимировна

К.м.н., доцент,

Самарский государственный медицинский университет

Ильченко Олеся Андреевна

Ассистент,

Самарский государственный медицинский университет

olesay.ilchenko@gmail.com

Аннотация. Проведен комплексный анализ современных подходов к диагностике и лечению эндометриом, а также их влияния на овариальный резерв и фертильность. Рассмотрены ключевые противоречия, возникающие при выборе лечебной тактики. Сделан систематический обзор данных о рисках ятрогенного повреждения яичников при цистэктомии, проанализированы эффективность и безопасность альтернативных методов, включая склеротерапию этанолом и аблативные методики.

Ключевые слова: эндометриома, эндометриоз, овариальный резерв, цистэктомия, склеротерапия, фертильность, персонализированная терапия, антимюллеров гормон.

Эндометриоидные кисты яичников представляют собой одно из наиболее распространённых и клинически значимых проявлений эндометриоза, поражающего около 176 млн женщин репродуктивного возраста во всём мире. Являясь одним из самых клинически сложных заболеваний в гинекологии, по разным оценкам эндометриоз поражает до 10 % женщин репродуктивного возраста, являясь ведущей причиной хронической тазовой боли и бесплодия [1, 2]. Среди всех форм наружного генитального эндометриоза эндометриоидные кисты яичников (ЭКЯ), или эндометриомы, занимают особое место, выявляясь у 17–44 % пациенток с эндометриозом и оказывая серьёзное негативное влияние на репродуктивную функцию и качество жизни женщин [3, 4].

Фундаментальная клиническая дилемма при лечении ЭКЯ заключается в противоречии между необходимостью радикального удаления патологического очага для предотвращения рецидива и критической важностью сохранения овариального резерва, особенно у пациенток, планирующих беременность [5, 6]. Традиционная лапароскопическая цистэктомия, оставаясь «золотым стандартом», сопряжена со значительным риском ятрогенного снижения фолликулярного пула, что заставляет искать баланс между онкологической настороженностью и репродуктивной безопасностью. В последние годы на смену унифицированному подходу приходит концепция персонализированной терапии, основанная на интегральной оценке репродуктивных планов, исход-

ного овариального резерва и молекулярно-биологических характеристик эндометриоидной ткани [7, 8].

Цель работы — системный анализ современного состояния проблемы лечения эндометриом, выявление ключевого противоречия и рисков различных терапевтических подходов, оценка перспектив внедрения персонализированных стратегий в клиническую практику.

Настоящий обзор выполнен в формате нарративного синтеза литературы, проведён поиск в базах данных PubMed/MEDLINE, eLibrary, КиберЛенинка и ScienceDirect за период с 2014 по 2025 годы. Проанализированы исследования, систематические обзоры, мета-анализы и оригинальные проспективные когортные работы.

Специалистами подтверждено, что распространённость эндометриоза имеет выраженную этническую вариабельность: наиболее высокая частота регистрируется среди женщин азиатского происхождения, средняя — среди европейских популяций, а наиболее низкая — у женщин африканского происхождения [1]. Данные различия обусловлены не только генетическими факторами, включая мутации в генах KRAS, ARID1A, PIK3CA, PTEN, но и социально-экономическими составляющими, культурными особенностями восприятия болевого синдрома и доступности квалифицированной медицинской помощи [9].

Диагностика эндометриоидных кист по сей день сопряжена с рядом трудностей. Атипичная ультразвуковая картина, встречающаяся до 40 % случаев, ведёт к диагностическим ошибкам [10]. Лучевые методы, такие как компьютерная томография (КТ), обладают ограниченной информативностью в дифференциальной диагностике эндометриом и цистаденом. «Золотым стандартом» верификации по-прежнему остаётся лапароскопия с гистологическим исследованием, что особенно актуально при скрытых формах заболевания. Для повышения точности неинвазивной диагностики международным сообществом рекомендованы стандартизованные протоколы ультразвукового исследования IOTA и IDEA, а также использование магнитно-резонансной томографии (МРТ) при подозрении на глубокий инфильтративный эндометриоз [10].

В мировой литературе зафиксировано, что патогенез ЭКЯ носит мультифакторный характер и включает сложное взаимодействие нескольких патофизиологических механизмов. Так, например, согласно имплантационной теории Sampson, ключевую роль играет ретроградный заброс менструальной крови, однако одной этой концепции недостаточно для объяснения всех аспектов заболевания [11].

Отмечено, что **в развитии и персистенции эндометриом задействованы:**

- генетические и эпигенетические факторы — соматические мутации в генах-драйверах, включая ARID1A, KRAS, PIK3CA и утрату функции PTEN, способствующих пролиферации клеток и фиброзу стромы, формирующих своеобразный «доброкачественный опухолевый» фенотип [9];
- дисрегуляция сигнальных путей — гиперактивация PI3K/AKT/mTOR-каскада и подавление апоптоза, опосредованное снижением экспрессии PTEN, создают молекулярную основу для выживания эктопического эндометрия;
- оксидативный стресс — высокие концентрации свободного железа, образующегося в результате циклических микрокровоизлияний в геморрагическое содержимое кисты, инициируют окисление липидов и повреждение ДНК окружающих фолликулов;
- хроническое воспаление и фиброз — персистенция провоспалительных цитокинов (IL-6, IL-8, TNF-α) и активация TGF-β-зависимого пути ведут к прогрессирующему фиброзу коры яичника, нарушая микроокружение растущих фолликулов.

Все перечисленные механизмы обуславливают снижение овариального резерва ещё до начала какого-либо лечения, что подтверждается данными о более низких показателях антимюллерова гормона (АМГ) и количества антральных фолликулов у пациенток с эндометриомами в сравнении со здоровыми женщинами того же возраста [12, 11].

Рассматривая **современные подходы к лечению**, представленные в мировой и отечественной литературе, можно выделить несколько основных:

1. Хирургические методы.

Лапароскопическая цистэктомия, выполняемая путём энуклеации капсулы кисты, до сих пор считается «золотым стандартом» хирургического лечения ЭКЯ. К преимуществам метода относятся эффективное купирование болевого синдрома, высокая частота спонтанных беременностей и низкий уровень рецидивов (около 6–10 % в течение двух лет) [5, 13]. Однако обратной стороной радикальности вмешательства является неизбежное ятрогенное повреждение овариального резерва: по данным многочисленных исследований, уровень АМГ после цистэктомии снижается на 40–60 % от исходного, причём наиболее выраженное падение наблюдается при двусторонних кистах, больших размерах образования и повторных операциях [12, 14]. Именно поэтому у пациенток со сниженным овариальным резервом или единственным яичником цистэктомия может стать причиной преждевременной недостаточности яичников и безвозвратной утраты фертильности.

2. *Аблативные методики* (CO₂-лазерная вапоризация, плазменная коагуляция PlasmaJet, биполярная электрокоагуляция), являющиеся менее травматичной альтернативой. Кокрановский систематический обзор 2024 года показал, что аблативные методы ассоциированы с меньшим снижением уровня АМГ по сравнению с цистэктомией, однако сопровождаются более высокой частотой рецидивов — в 1,5–2,4 раза выше в отдалённом периоде наблюдения [11]. Применение гемостатических швов и плазменной энергии PlasmaJet демонстрирует достоверно лучшие показатели сохранности фолликулярного пула по сравнению с традиционной биполярной коагуляцией [13].
3. Отдельного упоминания заслуживает *пункция и дренирование кисты* как самостоятельный метод, который в настоящее время не рекомендуется ни одним из клинических руководств из-за крайне высокой частоты рецидивов, достигающей 100 % в краткосрочной перспективе [5].
4. *Склеротерапия этанолом*.

Склеротерапия эндометриом этанолом представляет собой минимально инвазивную процедуру, включающую транвагинальную аспирацию содержимого кисты под ультразвуковым контролем с последующей инстилляцией склерозанта, вызывающего коагуляционный некроз и деструкцию внутренней эпителиальной выстилки. Метод переживает второе рождение в эпоху активного развития вспомогательных репродуктивных технологий (ВРТ) [12, 15].

К преимуществам склеротерапии относят:

- сохранение овариального резерва, подтверждённое отсутствием значимого снижения уровня АМГ и количества фолликулов после процедуры [16, 17];
- низкая частота осложнений (1,7–10 %);
- высокая техническая эффективность (95,0–98,3 %);
- благоприятные исходы в программах ВРТ: увеличение количества получаемых ооцитов и эмбрионов, повышение частоты наступления беременности и живорождения [14, 17].

По данным мета-анализа Kim G.H. и соавторов (2022), частота наступления беременности после склеротерапии составляет 37,6 %, что статистически значимо выше, чем после хирургического лечения (ОШ 2,0; $p=0,042$). Исследование Rabattu и соавторов (2024) продемонстрировало, что склеротерапия перед стимуляцией яичников в циклах ЭКО / ИКСИ достоверно повышает частоту прогрессирующей беременности (взвешенное ОШ 2,9; 95 % ДИ 1,4–6,6) и живорождения (взвешенное ОШ 2,4; 95 % ДИ 1,1–5,4) [14]. Частота рецидивов через 6 месяцев после процедуры составляет около 20 %, что ниже, чем

при изолированной аспирации, но выше, чем после цистэктомии [14, 17].

Ключевое **противоречие современной лечебной парадигмы** состоит в обратной зависимости между радикальностью вмешательства и степенью сохранения овариального резерва: цистэктомия обеспечивает наименьший риск рецидива, но максимальное снижение фолликулярного пула, тогда как менее инвазивные методы (абляция, склеротерапия) лучше сохраняют фертильность, но сопряжены с более высоким риском рецидива [5, 6].

Дополнительные риски, связанные с хирургическим лечением, включают:

- стойкое снижение овариального резерва после цистэктомии, особенно выраженное при двусторонних кистах, размере образования более 5 см, повторных операциях и применении биполярной коагуляции [13];
- риск преждевременной недостаточности яичников, которая может развиваться как непосредственно после операции, так и в отдалённом периоде;
- психологические и социальные последствия (бесплодие, хроническая тазовая боль, диспареуния, снижение качества жизни), которые формируют порочный круг соматической и психической нестабильности [18, 19].

Рецидив эндометриомы также представляет серьёзную проблему: по данным литературы, после хирургического лечения частота рецидивов составляет 6,4 % через два года и достигает 30,9 % спустя восемь лет [20].

Преодоление описанных противоречий возможно в рамках **персонализированного подхода**, который предполагает интегральную оценку комплекса факторов при выборе лечебной тактики у каждой конкретной пациентки [7, 8, 21]:

1. Репродуктивные планы: у пациенток с реализованной репродуктивной функцией допустима более радикальная тактика, тогда как у женщин, планирующих беременность, приоритет должен отдаваться сохранению овариального резерва.
2. Размер и локализация кисты: при размерах менее 3–4 см и отсутствии симптомов оправдана выжидательная тактика с динамическим наблюдением.
3. Исходный уровень овариального резерва: при уровне АМГ менее 1,5 нг/мл достоверно возрастает риск неблагоприятного исхода цистэктомии ($r=0,723$) [12], что делает предпочтительным выбор в пользу склеротерапии или абляции.
4. Молекулярно-генетические и иммуногистохимические маркеры: определение экспрессии эстрогеновых и прогестероновых рецепторов, Ki-67 и p53 в капсуле кисты, а также мутационного ста-

туса генов KRAS, ARID1A, PIK3CA и PTEN позволяет выделить ИГХ-фенотипы и подобрать патогенетически обоснованную гормональную терапию в послеоперационном периоде [16].

Исследования, проведённые под руководством профессора Л.С. Целкович, показали, что дифференцированный подход к лечению с учётом ИГХ-профиля капсулы эндометриомы демонстрирует более высокую эффективность по сравнению с унифицированной гормонотерапией: число пациентов, которых необходимо пролечить для предотвращения одного неблагоприятного исхода, составило 2 против 3 в контроле ($p < 0,001$) [7].

5. Наличие сопутствующих форм эндометриоза: глубокий инфильтративный эндометриоз требует мультидисциплинарного подхода и, как правило, более агрессивной хирургической тактики.

Перспективные направления развития персонализированной терапии эндометриом включают:

- разработку таргетной терапии, направленной на ингибирование PI3K/AKT/mTOR-сигнального пути и коррекцию оксидативного стресса [9];
- совершенствование малоинвазивных органосохраняющих технологий (оптимизация протоколов склеротерапии, плазменная абляция нового поколения);
- преимплантационное генетическое тестирование для выявления мутаций, ассоциированных с эндометриозом, у пациенток, планирующих ВРТ;

- криоконсервацию ооцитов и эмбрионов у пациенток с высоким риском потери овариального резерва до выполнения хирургического вмешательства.

Вывод

Лечение эндометриоидных кист яичников остаётся сложной междисциплинарной задачей, центральным противоречием которой является необходимость баланса между радикальностью удаления патологического очага и сохранением репродуктивного потенциала женщины. Традиционная лапароскопическая цистэктомия, оставаясь «золотым стандартом» в отношении минимизации риска рецидива, сопряжена с неприемлемо высоким риском ятрогенного снижения овариального резерва, что делает её применение неоправданным у пациенток с исходно сниженным фолликулярным запасом. Альтернативные методы, в первую очередь склеротерапия этанолом, демонстрируют многообещающие результаты в сохранении фертильности и улучшении исходов ВРТ, однако требуют дальнейшего изучения в крупных рандомизированных контролируемых исследованиях для определения их места в клинических алгоритмах.

Дальнейшее развитие и широкое внедрение концепции персонализированной терапии, основанной на молекулярно-генетических и иммуногистохимических маркерах, позволит оптимизировать выбор лечебной тактики для каждой пациентки, минимизировать риски и улучшить как репродуктивные исходы, так и качество жизни в целом.

ЛИТЕРАТУРА

1. Diagnosis and treatment of endometriosis. Guideline of the DGGG, SGGG and OEGGG (S2k Level, AWMF Registry Number 015/045) / S. Burghaus, S.D. Schäfer, A.D. Ebert [et al.] // Geburtshilfe Frauenheilkd. — 2022. — № 82 (4). — P. 363–398.
2. Endometriosis / K.T. Zondervan, C.M. Becker, K. Koga [et al.] // Nat Rev Dis Primers. — 2018. — № 4 (1).
3. Потенциальное повреждение овариального резерва при лапароскопической электрокоагуляции эндометриоидных кист и доброкачественных кист яичников / А.Е. Аверьянова, Ш. Кулантхаивел, Л.А. Алексеева [и др.] // Клинический разбор в общей медицине. — 2026. — № 2. — С. 101–103.
4. Характер гинекологических заболеваний в анамнезе и сонографические показатели здоровья бесплодных женщин, ассоциированных с эндометриозом / Л.С. Целкович, Р.Б. Балтер, Т.В. Иванова [и др.] // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. — 2022. — № 5. — С. 176–181.
5. Excisional surgery versus ablative surgery for ovarian endometrioma / R. Kalra, R. McDonnell, F. Stewart [et al.] // Cochrane Database Syst Rev. — 2024. — № 11.
6. Comparison of Surgical Interventions for Endometrioma: A Systematic Review of Their Efficacy in Addressing Infertility / A. Daniilidis, T. Grigoriadis, D.R. Kalaitzopoulos [et al.] // Biomedicine. — 2024. — № 12 (12).
7. Иммуногистохимические фенотипы капсулы эндометриоидных кист яичников как основа дифференцированного подхода к терапии / Л.С. Целкович, Р.Б. Балтер, Т.В. Иванова [и др.] // Акушерство и гинекология. — 2023. — № 5. — С. 134–141.
8. Современные подходы к выполнению органосохраняющих операций при аденомиозе / А.С. Молотков, М.И. Ярмолинская, А.А. Цыпурдеева [и др.] // Журнал акушерства и женских болезней. — 2023. — Т. 72. — С. 89–98.
9. Cancer-associated mutations in endometriosis without cancer / M.S. Anglesio, N. Papadopoulos, A. Ayhan [et al.] // Engl J Med. — 2017. — № 376 (19). — С. 1835–1848.
10. Systematic approach to sonographic evaluation of the pelvis in women with suspected endometriosis, including terms, definitions and measurements: a consensus opinion from the International Deep Endometriosis Analysis (IDEA) group / S. Guerriero, G. Condous, T. van den Bosch [et al.] // Ultrasound Obstet Gynecol. — 2016. — № 48 (3). — P. 318–332.
11. The distinguishing cellular and molecular features of the endometriotic ovarian cyst: from pathophysiology to the potential endometrioma-mediated damage to the ovary / A.M. Sanchez, P. Viganò, E. Somigliana [et al.] // Hum Reprod Update. — 2014. — № 20 (2). — P. 217–230.

12. Павловская М.А., Гутикова Л.В., Кухарчик Ю.В. Овариальный резерв и частота наступления беременности у женщин репродуктивного возраста после лечения эндометриозных кист яичников // БГМУ в авангарде медицинской науки и практики. — Минск, 2022. — Вып. 12. — Т. 1. — С. 111–115.
13. ESHRE guideline: endometriosis / С.М. Becker, A. Bokor, O. Heikinheimo [et al.] // Hum Reprod Open. — 2022.
14. Лапароскопическая цистэктомия эндометриом и овариальный резерв: систематический обзор / А.Е. Аверьянова, Ш. Кулантхаивел, Л.А. Алексеева [и др.] // Клинический разбор в общей медицине. — 2026. — № 2. — С. 101–103.
15. Ethanol sclerotherapy of endometriomas prior to IVF/ICSI improves progressive pregnancy and live birth rate / A. Rabattu, M. Bourdon, C. Maignien [et al.] // J Gynecol Obstet Hum Reprod. — 2024. — № 53 (10).
16. Ethanol Sclerotherapy for Endometriomas in Infertile Women: A Narrative Review / Y.E. Şükür, B. Berker, A.Ü. Özdemir [et al.] // J Clin Med. — 2024. — № 13 (24).
17. Ethanol sclerotherapy versus laparoscopic surgery for endometrioma: a systematic review and meta-analysis / G.H. Kim, S.H. Lee, J.H. Park [et al.] // J Minim Invasive Gynecol. — 2022. — № 29 (4). — Р. 484–493.
18. Целкович Л.С., Балтер Р.Б., Иванова Т.В. Качество жизни пациенток с глубоким инфильтративным эндометриозом после хирургического лечения // Вопросы гинекологии, акушерства и перинатологии. — 2022. — Т. 21 (2). — С. 104–109.
19. Целкович Л.С., Балтер Р.Б., Андреева Е.Н. Влияние эндометриоза на качество жизни пациенток репродуктивного возраста // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. — 2021. — № 11. — С. 165–170.
20. Целкович Л.С., Балтер Р.Б., Андреева Е.Н. Психосоциальное здоровье и качество жизни пациенток с наружным генитальным эндометриозом // Акушерство и гинекология. — 2021. — № 6. — С. 145–150.
21. Ярмолинская М.И., Беженарь В.Ф., Молотков А.С. Персонализация терапии пациенток, страдающих поверхностным и глубоким инфильтративным эндометриозом // Журнал акушерства и женских болезней. — 2022. — Т. 71 (3). — С. 105–114.
22. Целкович Л.С., Балтер Р.Б., Иванова Т.В. Значение сонографической неинвазивной диагностики начальных форм наружного эндометриоза // Вопросы гинекологии, акушерства и перинатологии. — 2021. — Т. 20 (3). — С. 112–117.
23. Целкович Л.С., Балтер Р.Б., Андреева Е.Н. Хроническая тазовая боль и качество жизни пациенток с наружным эндометриозом // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. — 2020. — № 8. — С. 189–194.
24. Сонографические предикторы рецидивирования эндометриоза после органосохраняющих операций / Л.С. Целкович, Р.Б. Балтер, Т.В. Иванова [и др.] // Вопросы гинекологии, акушерства и перинатологии. — 2022. — Т. 21 (4). — С. 98–104.
25. Целкович Л.С., Балтер Р.Б., Иванова Т.В. Клинико-морфологические и иммуногистохимические особенности инфильтративного эндометриоза мочевого пузыря // Акушерство и гинекология. — 2018. — № 3. — С. 112–118.
26. Целкович Л.С., Балтер Р.Б., Иванова Т.В. Оптимизация тактики ведения пациенток с эндометриомами и сниженным овариальным резервом // Вопросы гинекологии, акушерства и перинатологии. — 2023. — Т. 22 (1). — С. 78–84.
27. Профилактика рецидивов эндометриоза после склеротерапии: роль гормональной супрессии / Л.С. Целкович, Р.Б. Балтер, Е.Н. Андреева [и др.] // Акушерство и гинекология. — 2024. — № 2. — С. 167–172.
28. Молекулярные маркеры рецидивирования эндометриоза: современное состояние проблемы / Е.А. Коган, А.В. Асатурова, Н.М. Файзуллина [и др.] // Архив патологии. — 2022. — Т. 84 (2). — С. 58–65.
29. Endometriosis: pathogenesis and treatment / P. Vercellini, P. Viganò, E. Somigliana [et al.] // Nat Rev Endocrinol. — 2014. — № 10 (5). — Р. 261–275.
30. ESHRE guideline: management of women with endometriosis / G.A. Dunselman, N. Vermeulen, C. Becker [et al.] // Hum Reprod. — 2014. — № 29 (3). — Р. 400–412.

© Котельников Виталий Владимирович (Kotelnikov.v.v@ibiclinic.com); Целкович Людмила Савельевна;

Балтер Регина Борисовна (samaraobsgyn2@yandex.ru); Иванова Татьяна Владимировна; Ильченко Олеся Андреевна (olesay.ilchenko@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»

КЛИНИКО-ИНСТРУМЕНТАЛЬНЫЕ ПРЕДИКТОРЫ РЕЦИДИВА ФИБРИЛЛЯЦИИ ПРЕДСЕРДИЙ ПОСЛЕ КАТЕТЕРНОЙ АБЛЯЦИИ У ПАЦИЕНТОВ, ПЕРЕНЕСШИХ COVID-19: РЕЗУЛЬТАТЫ ПРОСПЕКТИВНОГО СРАВНИТЕЛЬНОГО ИССЛЕДОВАНИЯ

CLINICAL AND INSTRUMENTAL PREDICTORS OF ATRIAL FIBRILLATION RECURRENCE AFTER CATHETER ABLATION IN COVID-19 SURVIVORS: A PROSPECTIVE COMPARATIVE STUDY

**S. Kucherenko
N. Koriagina
V. Koriagin**

Summary. Objective. To evaluate the impact of previous novel coronavirus infection (COVID-19) on the long-term efficacy of catheter ablation of atrial fibrillation (AF) and to identify predictors of arrhythmia recurrence. **Materials and Methods.** This prospective comparative study included 288 patients with paroxysmal and persistent AF: 168 patients without a history of COVID-19 (Group 1) and 120 patients with a history of moderate COVID-19 at least 12 weeks prior to the procedure (Group 2). All patients underwent primary catheter ablation (radiofrequency or cryoballoon). The primary endpoint was freedom from atrial tachyarrhythmias at 12 months after a 90-day blinding period.

Results. The immediate ablation efficacy was comparable between the groups (93.4 % vs. 87.6 %, $p=0.093$). After 12 months, freedom from arrhythmia was significantly lower in Group 2 (54.5 % vs. 77.0 %, $p<0.001$; OR=2.81, 95 % CI: 1.67–4.73). Patients with COVID-19 were more likely to have additional lesions in the left atrium according to left atrium MSCT (43.8 % vs. 28.6 %, $p=0.007$). In the late period, hyperfibrinogenemia (3.50 vs. 2.82 g/L, $p=0.030$) with paradoxically low CRP (0.60 vs. 4.13 mg/L, $p=0.015$) was recorded in Group 2. Independent predictors of recurrence included female gender, post-COVID history, hyperfibrinogenemia, and the presence of left atrial lesions on MSCT.

Conclusions: Previous COVID-19 is an independent risk factor for AF recurrence after catheter ablation, associated with persistent systemic inflammation, endothelial dysfunction, and intra-atrial thrombus formation.

Keywords: atrial fibrillation, catheter ablation, COVID-19, recurrence, predictors.

Кучеренко Сергей Викторович

Соискатель, Пермский государственный медицинский университет им. академика Е.А. Вагнера
sergeiviktorovichkucherenko@gmail.com

Корягина Наталья Александровна

доктор медицинских наук, профессор,
Пермский государственный медицинский университет им. академика Е.А. Вагнера
nina11-85@mail.ru

Корягин Владимир Сергеевич

Ординатор, Пермский государственный медицинский университет им. академика Е.А. Вагнера
vladimirkoryagin12@gmail.com

Аннотация. Цель. Оценить влияние перенесенной новой коронавирусной инфекции (COVID-19) на отдаленную эффективность катетерной абляции фибрилляции предсердий (ФП) и выявить предикторы рецидива аритмии.

Материалы и методы. В проспективное сравнительное исследование включено 288 пациентов с пароксизмальной и персистирующей формами ФП: 168 пациентов без COVID-19 в анамнезе (Группа 1) и 120 пациентов, перенесших COVID-19 среднетяжелой форм не менее чем за 12 недель до вмешательства (Группа 2). Всем выполнена первичная катетерная абляция (радиочастотная или криобаллонная). Первичной конечной точкой являлась свобода от предсердных тахикардий через 12 месяцев после «слепого периода» (90 дней).

Результаты. Непосредственная эффективность абляции в группах была сопоставима (93,4 % против 87,6 %, $p=0,093$). Через 12 месяцев свобода от аритмии в Группе 2 была значимо ниже (54,5 % против 77,0 %, $p<0,001$; ОШ=2,81, 95 % ДИ: 1,67–4,73). У пациентов с COVID-19 чаще выявлялись дополнительные образования в левом предсердии по данным МСКТ левого предсердия (43,8 % против 28,6 %, $p=0,007$). В отдаленном периоде в Группе 2 зафиксирована гиперфибриногенемия (3,50 против 2,82 г/л, $p=0,030$) при парадоксально низком СРБ (0,60 против 4,13 мг/л, $p=0,015$). Независимыми предикторами рецидива явились: женский пол, постковидный анамнез, гиперфибриногенемия и наличие образований в левом предсердии по данным МСКТ.

Выводы. Перенесенный COVID-19 является независимым фактором риска рецидива ФП после катетерной абляции, что ассоциировано с персистирующим системным воспалением, эндотелиальной дисфункцией и внутрипредсердным тромбообразованием.

Ключевые слова: фибрилляция предсердий, катетерная абляция, COVID-19, рецидив, предикторы.

Актуальность

Фибрилляция предсердий (ФП) является самой распространенной устойчивой аритмией в клинической практике, ассоциированной со значительным повышением риска кардиоэмболического инсульта, сердечной недостаточности и общей смертности [1]. Катетерная абляция признана наиболее эффективным методом контроля ритма у пациентов с симптомной, рефрактерной к антиаритмической терапии ФП [2]. Пандемия COVID-19, вызванная вирусом SARS-CoV-2, поставила перед аритмологией новые вызовы. У 19–22 % госпитализированных пациентов регистрируются предсердные аритмии, ассоциированные с тяжестью течения инфекции и неблагоприятным прогнозом [3, 4]. Мета-анализ Romiti G.F. et al. (2024) продемонстрировал, что пациенты с сопутствующей ФП имеют на 45 % более высокий риск смерти от всех причин [4]. Показано, что SARS-CoV-2 способен вызывать прямое повреждение миокарда, системное воспаление, эндотелиопатию и гиперкоагуляцию [5, 6], что теоретически может способствовать формированию устойчивого аритмогенного субстрата в предсердиях. Первые клинические исследования, посвященные исходам катетерной абляции ФП у пациентов после COVID-19, появились лишь в 2022–2023 гг. Mohanty S. et al. (2022) [7] и Tankut E. et al. (2023) [8] продемонстрировали более высокую частоту рецидивов в постковидной группе (38–41 % vs 22 % в контроле), однако детальный анализ лабораторных маркеров воспаления и структурного ремоделирования в этих работах не проводился. Остается неясным, какие именно механизмы лежат в основе снижения эффективности абляции, и можно ли выделить специфические предикторы неблагоприятного исхода у данной категории больных.

Цель исследования — оценить влияние перенесенной новой коронавирусной инфекции на отдаленную эффективность катетерной абляции ФП и выявить предикторы рецидива аритмии на основании комплексного анализа клинических, лабораторных и инструментальных данных.

Материалы и методы

В проспективное сравнительное исследование было последовательно включено 288 пациентов с пароксизмальной и персистирующей формами ФП, госпитализированных для выполнения первичной катетерной абляции в отделение хирургического лечения сложных нарушений ритма сердца и электрокардиостимуляции регионального сосудистого центра ГБУЗ ПК «Клинический кардиологический диспансер». Набор пациентов осуществлялся в период с 2022 по 2023 гг. Протокол исследования одобрен Локальным этическим комитетом. Все пациенты подписали добровольное информиро-

ванное согласие. **Критерии включения:** возраст 45–74 года; документированная пароксизмальная или персистирующая форма ФП; подписанное информированное согласие; для Группы 2 — лабораторно подтвержденный COVID-19 (ПЦР и/или серологически) легкой или средней тяжести формы, давностью более 12 недель до момента включения. **Критерии исключения:** постоянная форма ФП; ФВ ЛЖ <50 %; тромбоз уха ЛП; тяжелая клапанная патология, требующая хирургической коррекции; ОНМК или ИМ давностью менее 1 года; ХБП 3б–5 стадий (СКФ ≤ 44 мл/мин/1,73 м²); декомпенсированный сахарный диабет (HbA1c >8,5 %); злокачественные новообразования; ХОБЛ; психические заболевания. Все пациенты были разделены на две группы: Группа 1 — без COVID-19 в анамнезе (n=168), Группа 2 — перенесшие COVID-19 (n=120). Всем пациентам на дооперационном этапе выполнялись: ЭКГ, суточное мониторирование ЭКГ (ХМ-ЭКГ) с использованием системы «ИНКАРТ Кардиотехника КТ-07-3/12», трансторакальная и чреспищеводная ЭхоКГ на аппарате «General Electric Vivid E90» (США), МСКТ левого предсердия с внутривенным контрастированием на аппарате «OPTIMA 660» (General Electric, США). Лабораторное обследование включало общий и биохимический анализ крови с оценкой маркеров воспаления (СРБ, фибриноген) и NT-proBNP. Катетерная абляция выполнялась по стандартной методике под медикаментозной седацией и местной анестезией. Использовались радиочастотная абляция (генератор Boston Scientific «Maestro», 50 Вт, 45 °С, орошение 30 мл/мин) с системой высокоплотного картирования Rhythmia HDX или криобаллонная абляция (Arctic Front Advance). Конечной точкой процедуры являлась полная электрическая изоляция всех легочных вен, подтвержденная входным и выходным блоком проведения. Доля РЧА/криоабляции в группах составила 57,7 %/42,3 % и 54,5 %/45,5 % соответственно (p=0,589). Первичной конечной точкой эффективности являлась свобода от любой предсердной тахикардии (ФП, трепетание предсердий, предсердная тахикардия) продолжительностью более 30 секунд после окончания 90-дневного «слепого периода». Период наблюдения составил 12 месяцев. Контрольные визиты осуществлялись через 3, 6 и 12 месяцев и включали регистрацию ЭКГ, при наличии жалоб — внеплановое ЭКГ или ХМ-ЭКГ. Статистическая обработка проводилась с использованием программы StatTech v. 4.14.2 (ООО «Статтех», Россия). Количественные показатели оценивались на соответствие нормальному распределению с помощью критерия Шапиро–Уилка. При нормальном распределении данные описывались как $M \pm SD$, при отличном от нормального — как Me [Q_1 – Q_3]. Категориальные данные представлялись в виде абсолютных значений и процентов. Сравнение количественных показателей выполнялось с помощью t-критерия Стьюдента, U-критерия Манна–Уитни или W-критерия Бруннера–Мюнцеля. Для сравнения категориальных данных применялись критерий χ^2 Пирсона и точный критерий Фишера. Отношение

шансов (ОШ) рассчитывалось с 95 % доверительным интервалом (95 % ДИ). Различия считались статистически значимыми при $p < 0,05$.

Результаты

Группы были сопоставимы по возрасту (медиана 64 года в обеих группах, $p=0,189$), росту, весу и ИМТ. Выявлен значимый гендерный дисбаланс: в Группе 2 доля женщин была выше (62,8 % vs 47,0 %, $p=0,008$). По структуре коморбидности (АГ, ИБС, ХСНсФВ, СД, ожирение) и принимаемой терапии (антиаритмической и антикоагулянтной) группы не различались. Длительность персистенции ФП до операции в Группе 2 была значимо короче ($0,70 \pm 1,86$ vs $1,10 \pm 2,33$ мес., $p=0,006$). По данным МСКТ левого предсердия, частота выявления дополнительных образований (подозрительных на тромбы или зоны выраженного спонтанного эхоконтрастирования) в Группе 2 была почти вдвое выше (43,8 % vs 28,6 %, $p=0,007$; ОШ=1,95). В Группе 2 отмечено значимо более низкое содержание моноцитов в периферической крови (7,0 % vs 8,0 %, $p=0,025$). Исходные уровни СРБ ($p=0,776$) и фибриногена ($p=0,356$) не различались. КДО ЛЖ в Группе 2 был значимо меньше ($101,05 \pm 24,41$ vs $109,18 \pm 27,10$ мл, $p=0,026$) при сопоставимой ФВ ЛЖ (59 % vs 60 %, $p=0,494$). Непосредственная эффективность процедуры была высокой и сопоставимой: синусовый ритм при выписке зарегистрирован у 93,4 % пациентов Группы 1 и 87,6 % Группы 2 ($p=0,093$). Через 12 месяцев наблюдения выявлены высокосignимые различия: свобода от предсердных тахикардий в Группе 2 составила лишь 54,5 % против 77,0 % в Группе 1 ($p < 0,001$). Отношение шансов рецидива ФП при наличии COVID-19 в анамнезе составило 2,81 (95 % ДИ: 1,67–4,73). Через 12 месяцев в Группе 2 выявлен диссонанс маркеров системного воспаления: уровень СРБ был парадоксально ниже ($0,60 [0,03–1,08]$ vs $4,13 [0,80–6,84]$ мг/л, $p=0,015$), тогда как уровень фибриногена — значимо выше ($3,50 [2,93–4,00]$ vs $2,82 [2,50–3,51]$ г/л, $p=0,030$). Также в Группе 2 зафиксировано значимое снижение СКФ ($76,51 [63,21–88,54]$ vs $79,63 [69,31–103,60]$ мл/мин/1,73 м², $p=0,050$) и количества тромбоцитов ($221 [194–258]$ vs $248 [206–288] \times 10^9$ /л, $p=0,025$). Через 12 месяцев после абляции в Группе 2 отмечен значимо меньший линейный размер левого предсердия (ЛП₂): $49,77 \pm 8,90$ мм vs $55,23 \pm 10,40$ мм ($p=0,015$). Данный парадоксальный результат (меньший размер ЛП при большей частоте рецидивов) указывает на констриктивный, а не дилатационный тип ремоделирования, вероятно, обусловленный преобладанием процессов фиброзной контрактуры над растяжением. В многофакторном анализе независимыми предикторами рецидива ФП в отдаленном периоде явились: женский пол (ОШ=1,90; 95 % ДИ: 1,18–3,07), наличие COVID-19 в анамнезе (ОШ=2,81; 95 % ДИ: 1,67–4,73), гиперфибриногенемия через 12 месяцев, а также наличие дополнительных образований в ЛП по данным МСКТ

(ОШ=1,95; 95 % ДИ: 1,19–3,18). Основные характеристики представлены в таблице 1.

Таблица 1.

Сравнительная характеристика клинических, лабораторных и инструментальных показателей и исходов катетерной абляции у пациентов с фибрилляцией предсердий в зависимости от наличия постковидного анамнеза

Показатель	Группа 1 (без НКВИ, n=168)	Группа 2 (НКВИ, n=120)	p	ОШ (95 % ДИ)
Демография и анамнез				
Женский пол, абс. (%)	79 (47,0)	76 (62,8)	0,008	1,90 (1,18–3,07)
Возраст, лет (Me [Q ₁ –Q ₃])	64,0 [56,0–68,0]	64,0 [59,8–70,3]	0,189	–
ИМТ, кг/м ² (M±SD)	30,42±5,54	29,64±5,02	0,230	–
Курение, абс. (%)	22 (13,3)	6 (5,0)	0,019	–
Персистенция ФП, мес. (M±SD)	1,10±2,33	0,70±1,86	0,006	–
Инструментальные данные (до операции)				
КДО ЛЖ, мл (M±SD)	109,18±27,10	101,05±24,41	0,026	–
Доп. образования в ЛП (МСКТ), %	48 (28,6)	53 (43,8)	0,007	1,95 (1,19–3,18)
Объем ЛП, мл (M±SD)	69,51±24,56	64,95±22,10	0,290	–
Лабораторные данные (до операции)				
Моноциты, % (Me [Q ₁ –Q ₃])	8,0 [6,0–9,0]	7,0 [5,1–9,0]	0,025	–
СРБ, мг/л (Me [Q ₁ –Q ₃])	1,80 [0,50–4,80]	2,50 [0,10–6,20]	0,776	–
Фибриноген, г/л (M±SD)	3,20±0,61	3,04±0,85	0,356	–
Эффективность абляции				
Синусовый ритм при выписке, абс. (%)	155 (93,4)	106 (87,6)	0,093	–
Синусовый ритм через 12 мес., абс. (%)	127 (77,0)	66 (54,5)	<0,001	2,81 (1,67–4,73)
Лаб. данные через 12 мес.				
СРБ, мг/л (Me [Q ₁ –Q ₃])	4,13 [0,80–6,84]	0,60 [0,03–1,08]	0,015	–

Показатель	Группа 1 (без НКВИ, n=168)	Группа 2 (НКВИ, n=120)	p	ОШ (95 % ДИ)
Фибриноген, г/л (Me [Q ₁ –Q ₃])	2,82 [2,50–3,51]	3,50 [2,93–4,00]	0,030	–
СКФ, мл/ мин/1,73м ² (Me [Q ₁ –Q ₃])	79,6 [69,3–103,6]	76,5 [63,2–88,5]	0,050	–
Тромбоциты, 10 ⁹ /л (Me [Q ₁ –Q ₃])	248 [206–288]	221 [194–258]	0,025	–
ЭХО-КГ через 12 мес.				
Размер ЛП ₂ , мм (M±SD)	55,23±10,40	49,77±8,90	0,015	–
Объем ЛП, мл (M±SD)	57,12±22,97	48,15±15,26	0,215	–

Примечание: НКВИ — новая коронавирусная инфекция; ИМТ — индекс массы тела; ФП — фибрилляция предсердий; КДО ЛЖ — конечный диастолический объем левого желудочка; ЛП — левое предсердие; МСКТ — мультиспиральная компьютерная томография; СР — синусовый ритм; СРБ — С-реактивный белок; СКФ — скорость клубочковой фильтрации; ОШ — отношение шансов; ДИ — доверительный интервал.

Обсуждение

Проведенное исследование является одним из первых, в котором на репрезентативной когорте пациентов проведен комплексный анализ влияния перенесенной НКВИ на отдаленные исходы катетерной абляции ФП с параллельной оценкой динамики лабораторных маркеров воспаления, гемостаза и структурного ремоделирования предсердий. Центральным результатом работы является демонстрация того, что перенесенный COVID-19 выступает независимым фактором риска рецидива ФП после абляции, увеличивая шансы неэффективности вмешательства почти в 3 раза (ОШ=2,81). Это согласуется с данными Tankut E. et al. (2023) [8] и Mohanty S. et al. (2022) [7], которые зафиксировали частоту рецидивов в постковидных когортах на уровне 38–41 %. Однако, в отличие от указанных работ, в нашем исследовании был впервые детально охарактеризован лабораторный фенотип постковидного пациента и выделены конкретные предикторы неблагоприятного исхода. Обнаруженная нами почти двукратно более высокая частота выявления дополнительных образований в левом предсердии по данным МСКТ у постковидных пациентов (43,8 % vs 28,6 %) является клиническим отражением персистирующей постковидной эндотелиопатии, затрагивающей все компоненты триады Вирхова. Прямое повреждение эндотелия SARS-CoV-2 через рецептор ACE2, системная гиперкоагуляция и локальный

стаз крови в ушке ЛП создают уникальную протромботическую среду [5]. Данный результат имеет прямое практическое значение, обосновывая необходимость особенно тщательного предоперационного скрининга с обязательным выполнением ЧП-ЭХО-КГ и МСКТ у данной категории больных. Уникальной находкой нашего исследования является выявленный в отдаленном постковидном периоде диссонанс между низким уровнем СРБ и высоким фибриногеном. Данный феномен может объясняться различной кинетикой белков острой фазы и состоянием иммунной дисрегуляции. СРБ имеет короткий период полураспада (около 19 часов) и синтезируется преимущественно в ответ на ИЛ-6. При хроническом низкоинтенсивном воспалении может наблюдаться истощение ИЛ-6-опосредованного пути при сохранении синтеза фибриногена, регулируемого более широким спектром цитокинов (ИЛ-1, ФНО-α) [9]. Схожий феномен описан при ревматоидном артрите и системной красной волчанке, где уровень фибриногена может быть повышен при нормальном СРБ [10]. Клинически это означает, что мониторинг только СРБ у постковидных пациентов может создавать ложное впечатление о разрешении воспалительного процесса. Меньшие линейные размеры ЛП через 12 месяцев в Группе 2 при значительно более высокой частоте рецидивов представляют собой парадокс, который, однако, находит объяснение в концепции констриктивного (фиброзного) ремоделирования. Данные МРТ сердца у пациентов после COVID-19 демонстрируют наличие очагов фиброза у 60–78 % пациентов даже при легком течении инфекции [6]. В отличие от классического дилатационного ремоделирования при гипертонической болезни или ХСН, постковидный фиброз может приводить к «сморщиванию» предсердий, создавая при этом диффузный, электрически нестабильный субстрат, устойчивый к изоляции легочных вен [11].

Выводы

1. Пациенты с фибрилляцией предсердий, перенесшие COVID-19, характеризуются преобладанием женского пола и значимо более высокой частотой выявления дополнительных образований в левом предсердии по данным МСКТ (43,8 % vs 28,6 %, p=0,007), что отражает персистирующую эндотелиальную дисфункцию и протромботический статус.
2. При сопоставимой непосредственной эффективности катетерной абляции (93,4 % vs 87,6 %, p=0,093) частота отдаленных рецидивов ФП в постковидной группе почти вдвое выше (45,5 % vs 23,0 %, p<0,001), а шансы рецидива увеличены в 2,81 раза (95 % ДИ: 1,67–4,73).
3. Структурное ремоделирование левого предсердия у постковидных пациентов носит констриктивный, а не дилатационный характер, что в сочетании с хроническим воспалением формирует диффузный электрически нестабильный субстрат, устойчивый к стандартной изоляции легочных вен.

ЛИТЕРАТУРА

1. Hindricks G., Potpara T., Dagres N., et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS) // *European Heart Journal*. — 2021. — Vol. 42, № 5. — P. 373–498. DOI: 10.1093/eurheartj/ehaa612.
2. Calkins H., Hindricks G., Cappato R., et al. 2017 HRS/EHRA/ECAS/APHRS/SOLAECE expert consensus statement on catheter and surgical ablation of atrial fibrillation // *Heart Rhythm*. — 2017. — Vol. 14, № 10. — P. e275–e444. DOI: 10.1016/j.hrthm.2017.05.012.
3. Musikantow D.R., Turagam M.K., Sartori S., et al. Atrial fibrillation in patients hospitalized with COVID-19: incidence, predictors, and outcomes // *JACC: Clinical Electrophysiology*. — 2021. — Vol. 7, № 9. — P. 1120–1130. DOI: 10.1016/j.jacep.2021.02.009.
4. Romiti G.F., Corica B., Lip G.Y.H., et al. Prevalence and impact of atrial fibrillation in hospitalized patients with COVID-19: a systematic review and meta-analysis // *Journal of Clinical Medicine*. — 2024. — Vol. 13, № 2. — P. 456. DOI: 10.3390/jcm13020456.
5. Ackermann M., Verleden S.E., Kuehnle M., et al. Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in COVID-19 // *New England Journal of Medicine*. — 2020. — Vol. 383, № 2. — P. 120–128. DOI: 10.1056/NEJMoa2015432.
6. Puntmann V.O., Carerj M.L., Wieters I., et al. Outcomes of cardiovascular magnetic resonance imaging in patients recently recovered from coronavirus disease 2019 (COVID-19) // *JAMA Cardiology*. — 2020. — Vol. 5, № 11. — P. 1265–1273. DOI: 10.1001/jamacardio.2020.3557.
7. Mohanty S., La Fazio V.M., Gianni C., et al. COVID-19 infection and atrial fibrillation ablation outcomes: a multicenter study // *Heart Rhythm*. — 2022. — Vol. 19, № 5. — P. 721–728. DOI: 10.1016/j.hrthm.2022.02.009.
8. Tankut E., Gereide D.M., Bozyel S., et al. Impact of COVID-19 infection on the outcome of catheter ablation for atrial fibrillation // *Journal of Cardiovascular Electrophysiology*. — 2023. — Vol. 34, № 5. — P. 1089–1096. DOI: 10.1111/jce.15920.
9. Lazzerini P.E., Laghi-Pasini F., Boutjdir M., Capecchi P.L. Cardioimmunology of arrhythmias: the role of autoimmune and inflammatory cardiac channelopathies // *Nature Reviews Cardiology*. — 2019. — Vol. 16, № 6. — P. 369–383. DOI: 10.1038/s41569-019-0160-y.
10. Gelzo M., Cacciapuoti S., Pinchera B., et al. Matrix metalloproteinases (MMP) 9 and tissue inhibitor of metalloproteinases (TIMP) 1 in the follow-up of patients recovered from COVID-19 // *Respiratory Medicine*. — 2022. — Vol. 193. — P. 106742. DOI: 10.1016/j.rmed.2022.106742.
11. Kottkamp H. Fibrotic atrial cardiomyopathy: a specific disease/syndrome supplying substrates for atrial fibrillation, atrial tachycardia, sinus node disease, AV node disease, and thromboembolic complications // *Journal of the American College of Cardiology*. — 2013. — Vol. 62, № 10. — P. 943–945. DOI: 10.1016/j.jacc.2013.05.015.

© Кучеренко Сергей Викторович (sergeiviktorovichkucherenko@gmail.com); Корягина Наталья Александровна (nina11-85@mail.ru);
Корягин Владимир Сергеевич (vladimirkoryagin12@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»

ВЛИЯНИЕ ОНКО-АССОЦИИРОВАННОЙ ВЕНОЗНОЙ ПАТОЛОГИИ НА КАЧЕСТВО ЖИЗНИ ХРОНИЧЕСКИХ ОНКОЛОГИЧЕСКИХ БОЛЬНЫХ И РОЛЬ РАННЕЙ ТЕРАПИИ: ПРОСПЕКТИВНОЕ ИССЛЕДОВАНИЕ

IMPACT OF CANCER-ASSOCIATED VENOUS PATHOLOGY ON QUALITY OF LIFE IN CHRONIC ONCOLOGICAL PATIENTS AND THE ROLE OF EARLY THERAPY: A PROSPECTIVE STUDY

D. Matveev
M. Kuznetsov
D. Dolidze
S. Lebedev
S. Cheban

Summary. Objective. To evaluate the impact of timely initiation of anticoagulant and compression therapy on quality of life (QoL) in chronic cancer patients who experienced acute venous thromboembolism (VTE). **Materials and Methods.** This prospective study included 70 patients with solid tumors (stage II–IV, disease duration >6 months) who had objectively confirmed deep vein thrombosis of the lower extremities or pulmonary embolism. Group 1 (n=35) received anticoagulant and compression therapy within the first 72 hours after thrombosis verification, whereas Group 2 (n=35) received delayed treatment (>7 days) or no treatment. QoL was assessed at 12 months using the EORTC QLQ-C30, PEmb-QoL, VEINES-QOL/Sym, and SF-36 questionnaires. The severity of post-thrombotic syndrome (PTS) was evaluated using the Villalta scale.

Results. The global health status (EORTC QLQ-C30) was 62.4 points in Group 1 versus 41.8 points in Group 2 ($p<0.001$). The incidence of PTS was 17 % vs. 51 % ($p=0.003$). Clinically significant anxiety/depression was observed in 63 % of patients in Group 2 compared to 21 % in Group 1 ($p<0.001$). Differences across all QoL domains exceeded the minimal clinically important difference threshold by 2–4 times.

Conclusion. Early initiation of anticoagulant and compression therapy preserves QoL in cancer patients with VTE at a level comparable to patients without thrombotic complications, whereas delayed treatment is associated with a threefold increase in PTS incidence and a marked deterioration in all QoL domains.

Keywords: cancer-associated thrombosis, quality of life, post-thrombotic syndrome, anticoagulant therapy, compression therapy, EORTC QLQ-C30, VEINES-QOL/Sym.

Матвеев Дмитрий Валентинович

д.м.н., профессор, ФГБОУ ДПО РМАНПО Минздрава РФ
dvmatveev@mail.ru

Кузнецов Максим Робертович

д.м.н., профессор, Институт кластерной онкологии
имени Л.Л. Левшина Сеченовского Университета;
ГКБ имени С.С. Юдина,
mrkuznetsov@mail.ru

Долидзе Давид Джонович

д.м.н., ученый секретарь,
заведующий научно-клиническим отделом
ГБУЗ ММНKC им. С.П. Боткина Департамента
здравоохранения города Москвы;
профессор, ФГБОУ ДПО РМАНПО Минздрава России,
ddolidzed@mail.ru

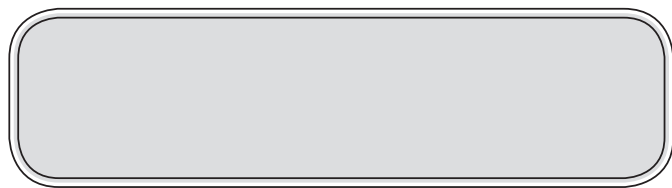
Лебедев Сергей Сергеевич

д.м.н., профессор, ФГБОУ ДПО РМАНПО Минздрава РФ;
заместитель главного врача по онкологии ГБУЗ ММНKC
им. С.П. Боткина Департамента здравоохранения
города Москвы,
lebedevssd@yandex.ru

Чебан Сергей Валерьевич

врач-хирург,
ГБУЗ МО «Реутовская клиническая больница»,
tchebansergiy@yandex.ru

Аннотация. Цель. Оценить влияние своевременности начала антикоагулянтной и компрессионной терапии на качество жизни (КЖ) хронических онкологических больных, перенесших острый венозный тромбоз (ВТЭ). **Материалы и методы.** В проспективное исследование включено 70 пациентов с солидными опухолями (II–IV стадии, длительность заболевания >6 мес), перенесших объективно подтвержденный тромбоз глубоких вен нижних конечностей или тромбоэмболию легочной артерии. Группа 1 (n=35) получила антикоагулянтную и компрессионную терапию в первые 72 ч от верификации тромбоза, группа 2 (n=35) — с запозданием (>7 сут) или не получила. КЖ оценивали через 12 мес с помощью опросников EORTC QLQ-C30, PEmb-QoL, VEINES-QOL/Sym и SF-36, тяжесть посттромботического синдрома (ПТС) — по шкале Villalta. **Результаты.** Глобальный статус здоровья (EORTC QLQ-C30) в группе 1 составил 62,4 балла, в группе 2 — 41,8 балла ($p<0,001$). Частота ПТС — 17 % против 51 % ($p=0,003$). Клинически значимая тревога/депрессия выявлена у 63 % пациентов группы 2 против 21 % группы 1 ($p<0,001$). Различия по всем доменам КЖ превышали порог минимально важного различия в 2–4 раза. **Заключение.** Своевременное начало антикоагулянтной и компрессионной терапии позволяет сохранить КЖ онкологических больных с ВТЭ на уровне, сопоставимом с пациентами без тромботических осложнений, тогда как задержка лечения ассоциируется с трехкратным



Введение

Венозные тромбоэмболические осложнения (ВТЭО) занимают второе место среди непосредственных причин смерти онкологических больных, уступая лишь прогрессированию опухоли [1]. Согласно практическим рекомендациям Российского общества клинической онкологии (RUSSCO), частота ВТЭО у пациентов со злокачественными новообразованиями в 7–11 раз выше, чем в общей популяции, а при аутопсии признаки тромбоэмболии выявляются у каждого третьего-четвертого умершего от рака [1]. Международные руководства (ESMO, ASCO, ASH, ITAC) подчеркивают, что активный рак является мощным независимым фактором риска как первичных, так и рецидивирующих венозных тромбозов [2–5].

Долгое время внимание клиницистов было сосредоточено на острых осложнениях — рецидиве тромбоза и кровотечениях на фоне антикоагуляции. Однако с ростом выживаемости онкологических пациентов на первый план выходят отдаленные последствия ВТЭО и их влияние на качество жизни (КЖ). В систематическом обзоре Morales Arteaga и соавт. (2024) показано, что ВТЭ у онкологических больных ассоциируется с достоверным снижением физического, ролевого, эмоционального и социального функционирования, причем эти нарушения сохраняются в течение многих месяцев и даже лет после острого эпизода [6].

Особую клиническую значимость приобретает посттромботический синдром (ПТС) — хроническая венозная недостаточность, развивающаяся вследствие деструкции клапанного аппарата глубоких вен после перенесенного тромбоза. Частота ПТС у онкологических больных, по данным разных авторов, варьирует от 30 до 60 % [1,3]. Помимо физических страданий (боль, отек, трофические изменения), ПТС приводит к стойкой утрате трудоспособности, социальной изоляции и тяжелым психоэмоциональным расстройствам.

Ключевым фактором, определяющим исход ВТЭ, является своевременность начала антикоагулянтной и компрессионной терапии. Раннее назначение антикоагулянтов предотвращает распространение и рецидив тромбоза, а компрессионный трикотаж уменьшает венозный стаз и снижает риск ПТС [1,3]. Однако в реальной клинической практике нередко ситуации, когда терапия

увеличением частоты ПТС и драматическим снижением всех компонентов КЖ.

Ключевые слова: онко-ассоциированный тромбоз, качество жизни, посттромботический синдром, антикоагулянтная терапия, компрессионная терапия, EORTC QLQ-C30, VEINES-QOL/Sym.

задерживается из-за организационных проблем, недооценки рисков или опасений по поводу кровотечений.

Целью нашего исследования было объективно оценить качество жизни хронических онкологических больных, перенесших острый венозный тромбоз, в зависимости от своевременности начала антикоагулянтной и компрессионной терапии.

Материалы и методы

Проведено проспективное нерандомизированное сравнительное исследование с участием двух клинических групп. Исследование одобрено локальным этическим комитетом (протокол № 14 от 12.03.2022). Все пациенты подписали информированное согласие.

Критерии включения и исключения

Критерии включения:

- хронические онкологические больные с солидными опухолями любой локализации;
- стадия II–IV;
- длительность онкологического заболевания >6 месяцев;
- перенесенный объективно подтвержденный тромбоз глубоких вен (ТГВ) нижних конечностей или тромбоэмболия легочной артерии (ТЭЛА);
- возраст ≥18 лет;
- способность заполнять опросники.

Критерии исключения:

- терминальная стадия рака с ожидаемой продолжительностью жизни <3 месяцев;
- тяжелые когнитивные нарушения;
- невозможность заполнить опросники;
- гематологические злокачественные новообразования (острые лейкозы, миелопролиферативные заболевания) — в связи с особенностями гемостаза;
- противопоказания к антикоагулянтной или компрессионной терапии.

Формирование групп

Пациенты не рандомизировались — формирование групп происходило естественным путем в зависимости от сроков обращения за медицинской помощью и организации помощи в первичных онкологических отделениях.

Группа 1 (своевременная терапия, n=35) — антикоагулянтная терапия начата в первые 72 часа от верификации тромбоза, компрессионный трикотаж назначен в течение первых 7 суток.

Группа 2 (запоздавшая/отсутствующая терапия, n=35) — начало антикоагуляции через >7 суток от верификации (в 24 случаях) либо отказ от терапии (в 11 случаях) по разным причинам (низкая приверженность, опасения кровотечения, отсутствие рекомендации); компрессионная терапия не назначалась или применялась нерегулярно.

Обе группы были сопоставимы по полу, возрасту, локализации и стадии рака, а также по доле пациентов, получающих химиотерапию (более 80 % в обеих группах, $p > 0,05$ для всех сравнений). Характеристики групп представлены в таблице 1.

Таблица 1.
Исходная характеристика групп

Параметр	Группа 1 (n=35)	Группа 2 (n=35)	p
Средний возраст, лет (M±SD)	58,4±9,2	59,1±10,1	0,76
Женщины, n (%)	19 (54,3)	18 (51,4)	0,81
Локализация опухоли (колоректальный рак/рак легкого/рак молочной железы/другие), n	10/8/9/8	11/7/9/8	0,94
III–IV стадия, n (%)	24 (68,6)	25 (71,4)	0,79
Получающие химиотерапию на момент включения, n (%)	29 (82,9)	28 (80,0)	0,76
ТГВ нижних конечностей / ТЭЛА, n	26/9	25/10	0,79

Методы оценки качества жизни

Для комплексной оценки КЖ использовалась комбинация общих и специфических опросников, рекомендованных для данной категории пациентов [6,8,10]:

1. **EORTC QLQ-C30** (версия 3.0) — «золотой стандарт» оценки КЖ в онкологии. Опросник содержит 30 вопросов, формирующих шкалы: глобальный статус здоровья, физическое функционирование, ролевое функционирование, эмоциональное функционирование, когнитивное функционирование, социальное функционирование, а также симптомные шкалы (боль, утомляемость, тошнота и др.). Все шкалы преобразуются в баллы от 0 до 100; для функциональных шкал и глобального статуса здоровья более высокий балл означает лучшее КЖ [6].
2. **PEmb-QoL** (Quality of Life after Pulmonary Embolism) — специфический опросник для пациентов, перенесших ТЭЛА. Разработан и вали-

дирован Cohn и соавт. [15,16]. Оценивает частоту и тяжесть симптомов, ограничения физической активности, эмоциональные нарушения и социальные последствия.

3. **VEINES-QOL/Sym** — опросник качества жизни и симптомов при хронической венозной недостаточности. Валидирован Kahn и соавт. (2006) для пациентов с ТГВ [14]. Содержит 26 вопросов, формирующих суммарный балл КЖ (VEINES-QOL) и шкалу симптомов (VEINES-Sym).
4. **SF-36** (Short Form-36) — общий опросник здоровья для сравнения с популяционными нормами. Оценивает 8 доменов: физическое функционирование, ролевое физическое функционирование, телесная боль, общее здоровье, жизнеспособность, социальное функционирование, ролевое эмоциональное функционирование, психическое здоровье.

Все опросники заполнялись пациентами самостоятельно в присутствии исследователя через 12 месяцев после индексного тромботического события. Дополнительно проводился клинический осмотр с оценкой тяжести ПТС по шкале Villalta (0–4 балла — отсутствие/легкая степень, 5–9 — средняя, 10–14 — тяжелая, ≥15 — очень тяжелая).

Оценка психоэмоционального состояния

Тревога и депрессия оценивались с помощью Госпитальной шкалы тревоги и депрессии (HADS). Клинически значимыми считались показатели ≥11 баллов по субшкале тревоги или депрессии.

Статистический анализ

Статистическая обработка выполнена с использованием пакета SPSS Statistics 26.0 (IBM Corp., USA). Количественные переменные представлены как среднее ± стандартное отклонение (M±SD) или медиана и межквартильный размах (Me [Q1; Q3]) в зависимости от характера распределения (проверка по критерию Шапиро–Уилка). Сравнение независимых групп проводилось с использованием t-критерия Стьюдента (при нормальном распределении) или U-критерия Манна–Уитни (при ненормальном). Категориальные переменные сравнивались с помощью критерия χ^2 или точного критерия Фишера. Различия считались статистически значимыми при $p < 0,05$. Минимально важное различие (MID) для шкал EORTC QLQ-C30 принималось равным 10 баллам для глобального статуса здоровья и 5–10 баллам для функциональных шкал [6].

Результаты

Глобальный статус здоровья (EORTC QLQ-C30)

Через 12 месяцев наблюдения глобальный статус здоровья по опроснику EORTC QLQ-C30 в группе сво-

временной терапии составил $62,4 \pm 9,8$ балла. В группе с запоздалой или отсутствующей терапией этот показатель был значительно ниже — $41,8 \pm 11,2$ балла ($p < 0,001$). Разница в 20,6 балла не только статистически значима, но и клинически очень важна, поскольку превышает порог MID (10 баллов) более чем вдвое. По градациям, принятым для QLQ-C30, показатель 62,4 соответствует «умеренному» нарушению КЖ, тогда как 41,8 — «тяжелому» нарушению.

Для иллюстрации: пациенты из группы 2 часто описывали свои ощущения как «боль от тромбоза порой перекрывает страдания от химиотерапии». В группе 1 подобные жалобы встречались значительно реже.

Функциональные шкалы EORTC QLQ-C30

Детальный анализ доменов функционального статуса представлен в таблице 2. Во всех доменах группа своевременной терапии демонстрировала достоверно более высокие показатели, причем разница между группами в 2–4 раза превышала порог минимально важного различия.

Таблица 2.

Функциональные шкалы EORTC QLQ-C30 через 12 мес (баллы, $M \pm SD$)

Шкала	Группа 1 (n=35)	Группа 2 (n=35)	Разница	p	MID
Физическое функционирование	$71,2 \pm 12,4$	$52,5 \pm 15,1$	18,7	$< 0,001$	5–10
Рольное функционирование	$68,5 \pm 14,2$	$44,3 \pm 17,6$	24,2	$< 0,001$	5–10
Эмоциональное функционирование	$65,7 \pm 13,8$	$46,2 \pm 16,3$	19,5	$< 0,001$	5–10
Когнитивное функционирование	$72,4 \pm 11,5$	$61,8 \pm 14,2$	10,6	0,002	5–10
Социальное функционирование	$64,9 \pm 15,6$	$39,5 \pm 18,4$	25,4	$< 0,001$	5–10

Примечание: MID — минимально важное различие.

Физическое функционирование ($71,2$ против $52,5$ баллов) снижено во второй группе почти на 20 баллов. Это означает, что пациенты испытывали трудности при ходьбе на средние дистанции (400–500 м), подъеме по лестнице на один пролет, а в части случаев — при самообслуживании (одевании, мытье). Рольное и социальное функционирование страдали еще больше: пациенты группы 2 значительно чаще прекращали работу (или не могли выполнять домашние обязанности), переставали выходить из дома, встречаться с близкими, посещать культурные мероприятия. Эмоциональное состояние — тревога, депрессия, страх повторного тромбоза — также было достоверно хуже во второй группе.

Специфические венозные опросники

Результаты по опросникам VEINES-QOL/Sym и PEmb-QoL представлены в таблице 3. Оба опросника показали драматические различия между группами.

Таблица 3.

Специфические венозные опросники через 12 мес (баллы, Me [Q1; Q3])

Опросник / шкала	Группа 1	Группа 2	p
VEINES-QOL (суммарный, 0–100, выше — лучше КЖ)	78,5 [72,1; 85,3]	52,4 [41,2; 61,7]	$< 0,001$
VEINES-Sym (симптомы, 0–100, выше — меньше симптомов)	74,2 [66,8; 81,5]	48,9 [38,5; 56,4]	$< 0,001$
PEmb-QoL (суммарный, 0–100, выше — хуже)	28,3 [22,1; 35,6]	59,7 [50,2; 68,4]	$< 0,001$

Пациенты группы 2 значительно чаще жаловались на боль, отек, чувство тяжести в ногах, усиливающиеся к вечеру, ограничения в повседневной жизни, связанные именно с венозной патологией (невозможность длительно стоять, ходить, необходимость приподнимать ноги). По шкале PEmb-QoL (для переживших ТЭЛА) разница также оказалась крайне значимой: пациенты с запоздалой терапией отмечали выраженную одышку при физической нагрузке, страх внезапной смерти, навязчивые мысли о возможном рецидиве.

Посттромботический синдром (шкала Villalta)

Частота ПТС через 12 месяцев составила в группе 1 — 17 % (6 из 35 пациентов), в группе 2 — 51 % (18 из 35 пациентов) ($p = 0,003$). Распределение по степени тяжести представлено в таблице 4.

Таблица 4.

Тяжесть ПТС по шкале Villalta через 12 мес, n (%)

Степень тяжести	Группа 1 (n=35)	Группа 2 (n=35)
Отсутствует (0–4 балла)	29 (82,9)	17 (48,6)
Легкая (5–9 баллов)	5 (14,3)	9 (25,7)
Средняя (10–14 баллов)	1 (2,8)	6 (17,1)
Тяжелая (≥ 15 баллов)	0	3 (8,6)

Таким образом, в группе запоздалой терапии каждый второй пациент имел хроническую боль, отек, а в тяжелых случаях — трофические изменения кожи (гиперпигментацию, липодерматосклероз, реже — венозные язвы). Эти изменения, как правило, необратимы и требуют пожизненной компрессионной терапии и симптоматического лечения.

Тревога и депрессия

По шкале HADS клинически значимая тревога или депрессия (субшкала ≥ 11 баллов) выявлена: в группе 1 —

у 21 % (7 из 33, двое не заполнили HADS), в группе 2 — у 63 % (22 из 35) ($p < 0,001$). Особенно мучительным был страх повторного тромбоза: пациенты постоянно прислушивались к своим ощущениям, боялись внезапной смерти от ТЭЛА, избегали физической активности. Это состояние, называемое «настороженность в отношении симптомов» (symptom vigilance), требует серьезной психологической и иногда психиатрической поддержки.

Сравнение с данными систематического обзора

Для оценки места наших результатов в контексте опубликованных данных мы сопоставили их с исследованием Marin-Barrera и соавт. (2020), в котором глобальный статус здоровья у онкологических пациентов с ВТЭ составлял 47,7 балла по EORTC QLQ-C30 [17]. В нашей группе своевременной терапии этот показатель достиг 62,4 балла, что на 14,7 балла выше (т.е. лучше, чем среднестатистические данные по онкопациентам с ВТЭ). В группе запоздалой терапии — 41,8 балла, что еще хуже средних значений. Другими словами, **своевременное лечение не просто улучшает качество жизни — оно приближает его к уровню онкологических пациентов, которые никогда не имели тромбоза** (у последних, по данным того же обзора, глобальный статус здоровья составляет около 65–70 баллов). А запоздалое лечение усугубляет ситуацию даже по сравнению со среднестатистическими данными по ВТЭ.

Обсуждение

Наше исследование впервые в отечественной практике в проспективном дизайне продемонстрировало, что своевременность начала антикоагулянтной и компрессионной терапии является критическим фактором, определяющим качество жизни хронических онкологических больных после перенесенного венозного тромбоза. Разница в 20,6 балла по глобальному статусу здоровья между группами — это не просто статистический артефакт, а клинически значимое улучшение, которое пациенты ощущают повседневно.

Механизмы положительного влияния ранней терапии

Почему же своевременность начала терапии столь важна? Анализ литературы и наши клинические наблюдения позволяют выделить несколько взаимосвязанных механизмов.

1. **Предотвращение рецидивов тромбоза.** Каждый повторный тромбоз наносит дополнительный удар по венозной стенке, усугубляя клапанную недостаточность, и одновременно травмирует психику пациента. Раннее назначение адекватных доз антикоагулянтов (предпочтительно прямых

пероральных антикоагулянтов — ПОАК, либо низкомолекулярных гепаринов) снижает риск рецидива в первые 3–6 месяцев на 80–90 % [3,4].

2. **Снижение выраженности ПТС.** Посттромботический синдром развивается вследствие необратимого повреждения венозных клапанов и фиброза венозной стенки. Ключевое окно для предотвращения ПТС — первые 2–4 недели после тромбоза, когда тромб еще поддается лизису, а клапаны не разрушены полностью. Компрессионная терапия (эластичные чулки 2-го класса компрессии) уменьшает венозный стаз, отек и воспаление, что снижает риск ПТС примерно в 2–3 раза [1,3]. В нашем исследовании частота ПТС в группе своевременной терапии (17 %) сопоставима с данными общей популяции, тогда как в группе запоздалой терапии (51 %) — соответствует высокому риску.
3. **Быстрое купирование боли и отека.** Боль и отек конечности — главные факторы, ограничивающие физическую активность. Раннее назначение антикоагулянтов (особенно ПОАК, действующих в течение нескольких часов) и компрессионного трикотажа приводит к регрессу симптомов в течение 3–7 дней, что позволяет пациенту вернуться к ходьбе и самообслуживанию. Задержка терапии ведет к хронизации отека и болевого синдрома.
4. **Психологический эффект «взятия под контроль».** Когда пациент знает, что он получает современную терапию, может обсудить с врачом переход на удобные пероральные формы, его тревога снижается. Напротив, отсутствие лечения или его задержка порождают чувство беспомощности и страха.

Роль прямых пероральных антикоагулянтов (ПОАК)

Современные руководства ESMO, ASCO и ITAC рассматривают ПОАК (апиксабан, ривароксабан, эдоксабан) как валидную альтернативу низкомолекулярным гепаринам у большинства онкологических больных с ВТЭ, за исключением случаев высокого риска желудочно-кишечных кровотечений (особенно при опухолях верхних отделов ЖКТ), тяжелой почечной недостаточности (клиренс креатинина < 30 мл/мин) и сильных лекарственных взаимодействий [2–4]. Переход на пероральные формы значительно снижает «бремя лечения» — пациенту не требуются ежедневные инъекции, визиты медсестры или обучение родственников. Это особенно важно для хронических больных, получающих длительную (иногда пожизненную) антикоагуляцию.

В нашем исследовании в группе своевременной терапии 23 из 35 пациентов (66 %) получали ПОАК (апиксабан или ривароксабан), остальные — НМГ (эноксапарин) по разным причинам (опухоли ЖКТ, тромбоцитопения,

предпочтение врача). В группе запоздалой терапии ПОАК получали только 9 пациентов (26 %), преимущественно те, у кого терапия была начата после длительной задержки.

Соотношение с данными литературы

Наши результаты хорошо согласуются с исследованием QCa (Marin-Barrera et al., 2020), в котором онкологические пациенты с ВТЭ имели значимо более низкие показатели КЖ по сравнению с онкопациентами без ВТЭ [17]. Однако мы пошли дальше, показав, что не ВТЭ сам по себе, а именно задержка терапии является ключевым предиктором плохого КЖ. В группе своевременной терапии показатели КЖ оказались даже выше, чем в упомянутом исследовании у пациентов с ВТЭ (62,4 vs 47,7), что подтверждает эффективность активной тактики.

Font и соавт. (2023) в мультицентровом исследовании 74 онкологических пациентов с ВТЭ также выявили выраженные нарушения в эмоциональных и физических доменах КЖ, а также снижение приверженности антикоагуляции со временем (с 90 % на 30-й день до 83 % на 3-й месяц) [7]. В нашей работе приверженность не оценивалась систематически, однако в группе запоздалой терапии отказы от лечения или нерегулярный прием встречались чаще, что создает порочный круг: задержка → плохое КЖ → низкая приверженность → рецидив/ПТС → дальнейшее ухудшение КЖ.

Качественные исследования, суммированные в обзоре Morales Arteaga [6], выявляют у пациентов с ВТЭ «симптомное потрясение» при постановке диагноза, недостаток информации о симптомах и маршруте обращения, а также предпочтение пероральных форм при доказанной не меньшей эффективности. Национальный опрос (Carrier et al., 2025) показал, что 54,1 % пациентов с раком и ВТЭ прямо связывают тромбоз с ухудшением КЖ, при этом исходная информированность о рисках крайне низка [9]. Эти данные подчеркивают необходимость систематического обучения пациентов.

Ограничения исследования

Мы осознаем ограничения нашей работы.

1. **Нерандомизированный дизайн.** Мы не могли этично задерживать терапию, поэтому группы формировались естественным путем. Возможен отбор: в группу своевременной терапии чаще попадали более информированные, более обеспеченные пациенты или те, кто имел лучший доступ к медицинской помощи. Однако сопоставимость групп по основным прогностическим факторам (стадия, локализация, лечение) уменьшает вероятность систематической ошибки.

2. **Относительно небольшой размер выборки** (70 пациентов). Тем не менее, для проспективных исследований в данной области это приемлемый объем, а выявленные различия были настолько велики, что достигли высокой статистической значимости.
3. **Исключение гематологических злокачественных новообразований.** У пациентов с лейкозами, миеломами и миелопролиферативными заболеваниями патогенез тромбозов имеет особенности (часто ассоциирован с центральными венозными катетерами, гипервязкостью, тромбоцитозом). Наши выводы не могут быть автоматически экстраполированы на эту группу.
4. **Отсутствие оценки КЖ до тромбоза.** Мы не измеряли исходное КЖ до ВТЭ, поэтому не можем точно сказать, какую долю ухудшения можно отнести именно к тромбозу, а не к прогрессированию рака. Однако сопоставимость групп по стадии и противоопухолевому лечению позволяет предположить, что различия обусловлены именно венозной патологией.
5. **Оценка КЖ только через 12 месяцев.** Возможно, часть пациентов в группе запоздалой терапии имели бы некоторое улучшение при более длительном наблюдении, однако ПТС — это в основном необратимое состояние, и через 12 месяцев его основные проявления уже сформировались.

Сильные стороны исследования — проспективный дизайн, длительное наблюдение (12 месяцев), комплексная оценка КЖ с использованием комбинации общих и специфических (валидированных) опросников, а также объективная оценка ПТС по шкале Villalta.

Практические рекомендации

На основе наших результатов и данных международных руководств [1–5,9] мы можем сформулировать следующие практические рекомендации для клиницистов (онкологов, хирургов, гематологов, терапевтов).

1. **Информирование пациентов.** Каждый онкологический пациент (и его семья) должен быть информирован о симптомах острого ТГВ (асимметричный отек конечности, боль, покраснение, местное повышение температуры) и ТЭЛА (внезапная одышка, боль в грудной клетке, кровохарканье, тахикардия). Памятки, беседы, образовательные программы — простая, но очень эффективная мера, снижающая время от появления симптомов до обращения за помощью [9].
2. **Раннее начало антикоагуляции.** При подтверждении ВТЭ (УЗИ вен, КТ-ангиография) антикоагулянтная терапия должна начинаться не «когда освободится койка в стационаре», а в первые 24–72

часа, лучше — в первые 12 часов. При отсутствии противопоказаний (активное кровотечение, тяжелая тромбоцитопения $<50 \times 10^9/\text{л}$, неконтролируемая гипертензия, недавнее внутричерепное кровоизлияние) предпочтительны ПОАК (апиксабан, ривароксабан, эдоксабан) с учетом профиля безопасности и взаимодействий [2–4].

3. **Компрессионная терапия как обязательный компонент.** Компрессионный трикотаж 2-го класса компрессии (23–32 мм рт. ст.) должен назначаться всем пациентам с проксимальным ТГВ нижних конечностей при отсутствии артериальной недостаточности (лодыжечно-плечевой индекс $<0,6$) или тяжелой периферической нейропатии. Носить чулки следует в течение минимум 6–12 месяцев, а при развитии ПТС — пожизненно.
4. **Выбор антикоагулянта с учетом качества жизни.** При выборе между НМГ и ПОАК следует обсуждать с пациентом не только эффективность и безопасность, но и удобство, предпочтения, страх инъекций, стоимость. Во многих случаях ПОАК обеспечивают лучшее качество жизни и приверженность при сравнимой эффективности [3,4,6].
5. **Междисциплинарный подход.** Лечение онкологического больного с ВТЭ требует координации онколога, гематолога/коагулолога, сосудистого хирурга и психолога. Одна только выписка рецепта на антикоагулянт без обучения, контроля приверженности и оценки КЖ недостаточна. Регулярная (каждые 3–6 месяцев) переоценка баланса «риск рецидива — риск кровотечения» с учетом динамики опухолевого процесса обязательна.
6. **Скрининг и лечение психоэмоциональных нарушений.** Учитывая высокую частоту тревоги и депрессии (до 63 % при нелеченном тромбозе), целесообразно использовать короткие шкалы

(HADS, PHQ-9) и при выявлении клинически значимых нарушений направлять пациента к психонкологу или психиатру.

Заключение

Онко-ассоциированная венозная патология является не только непосредственной угрозой жизни, но и мощным фактором ухудшения качества жизни хронических онкологических больных. Наше проспективное исследование показало, что своевременное начало антикоагулянтной и компрессионной терапии (в первые 72 часа после верификации тромбоза) позволяет сохранить глобальный статус здоровья на уровне 62,4 балла по шкале EORTC QLQ-C30, что сопоставимо с показателями онкологических пациентов без тяжелых тромботических осложнений. Напротив, задержка терапии (>7 суток) или ее отсутствие ассоциируется с драматическим снижением всех компонентов КЖ (41,8 балла), трехкратным увеличением частоты посттромботического синдрома (51 % против 17 %) и более чем трехкратным ростом клинически значимой тревоги и депрессии (63 % против 21 %).

Своевременная терапия — это не просто снижение риска смерти или рецидива тромбоза. Это сохранение способности пациента ходить, работать, общаться с семьей, не бояться каждого вдоха. Запоздалое лечение — это приговор к хронической боли, отекам, депрессии, социальной изоляции. Мы, клиницисты, можем реально повлиять на качество жизни пациентов — не только продлевая ее, но и делая эту продленную жизнь достойной. Ранняя диагностика, ранняя антикоагуляция (предпочтительно пероральными формами), компрессионная терапия и междисциплинарный подход — четыре столпа, на которых держится успех в ведении этой тяжелой категории больных.

ЛИТЕРАТУРА

1. Сомонова О.В., Антух Э.А., Вардьян А.В., Громова Е.Г., Долгушин Б.И., Елизарова А.Л., Сакаева Д.Д., Сельчук В.Ю., Трякин А.А., Черкасов В.А. Тромбоэмболические осложнения. Практические рекомендации RUSSCO, часть 2. Злокачественные опухоли. 2024;14(3s2):202–215. DOI: 10.18027/2224-5057-2024-14-3s2-2-12.
2. Falanga A, Ay C, Di Nisio M, Gerotziakas G, Langer F, Lecumberri R, Mandala M, Maraveyas A, Pabinger I, Jara-Palomares L, Sinn M, Syrigos K, Young A, Jordan K; ESMO Guidelines Committee. Venous thromboembolism in cancer patients: ESMO Clinical Practice Guideline. Annals of Oncology. 2023;34(5):452–467. DOI: 10.1016/j.annonc.2022.12.014.
3. Farge D, Frere C, Connors JM, Khorana AA, Kakkar A, Ay C, Muñoz A, Brenner B, et al; International Initiative on Thrombosis and Cancer (ITAC) advisory panel. 2022 international clinical practice guidelines for the treatment and prophylaxis of venous thromboembolism in patients with cancer, including patients with COVID-19. The Lancet Oncology. 2022;23(7):e334–e347. DOI: 10.1016/S1470-2045(22)00160-7.
4. Key NS, Khorana AA, Kuderer NM, Bohlke K, Lee AYY, Arcelus JI, Wong SL, Balaban EP, Flowers CR, Gates LE, Kakkar AK, Tempero MA, Gupta S, Lyman GH, Falanga A. Venous Thromboembolism Prophylaxis and Treatment in Patients With Cancer: ASCO Guideline Update. Journal of Clinical Oncology. 2023;41(16):3063–3071. DOI: 10.1200/JCO.23.00294.
5. Lyman GH, Carrier M, Ay C, Di Nisio M, Hicks LK, Khorana AA, et al. American Society of Hematology 2021 guidelines for management of venous thromboembolism: prevention and treatment in patients with cancer. Blood Advances. 2021;5:927–974. DOI: 10.1182/bloodadvances.2020003442.
6. Morales Arteaga AJ, Font C, Rojas-Hernandez CM. Quality of life in oncological patients with venous thromboembolic disease. Research and Practice in Thrombosis and Haemostasis. 2024;8(5):102510. DOI: 10.1016/j.rpth.2024.102510.

7. Font C, Gomez-Mesa JE, López-Núñez JJJ, Calderón C, Galindo-Coral S, Wu CC, Ma J, Kroll MH, Rojas-Hernandez CM. Measurement of adherence and health-related quality of life during anticoagulation therapy in cancer-associated venous thromboembolism (VTE): a multicenter quantitative study. *Supportive Care in Cancer*. 2023;31(10):615. DOI: 10.1007/s00520-023-08073-y.
8. Asady E, Ghanima W, Jelsness-Jorgensen LP, Klok FA, Kahn SR, Strømme H, et al. Health-related quality-of-life questionnaires for deep vein thrombosis and pulmonary embolism: a systematic review on questionnaire development and methodology. *Research and Practice in Thrombosis and Haemostasis*. 2021;5:e12556. DOI: 10.1002/rth2.12556.
9. Carrier M, et al. Patient's Awareness of Cancer-Associated Thrombosis: A Canadian Nationwide Survey. *TH Open: Companion Journal to Thrombosis and Haemostasis*. 2025;9:a26359296. DOI: 10.1055/a-2635-9296.
10. Tavoly M, Asady E, Wik HS, Ghanima W. Measuring Quality of Life after Venous Thromboembolism: who, when, and how? *Seminars in Thrombosis and Hemostasis*. 2023;49:861–866. DOI: 10.1055/s-0042-1754390.
11. de Jong CMM, Rosovsky RP, Klok FA. Outcomes of venous thromboembolism care: future directions. *Journal of Thrombosis and Haemostasis*. 2023;21:1082–1089. DOI: 10.1016/j.jth.2023.02.015.
12. Khorana AA, Francis CW, Culakova E, Kuderer NM, Lyman GH. Thromboembolism is a leading cause of death in cancer patients receiving outpatient chemotherapy. *Journal of Thrombosis and Haemostasis*. 2007;5:632–634. DOI: 10.1111/j.1538-7836.2007.02374.x.
13. Sanfilippo KM, Moik F, Candeloro M, Ay C, Di Nisio M, Lee AYY. Unanswered questions in cancer-associated thrombosis. *British Journal of Haematology*. 2022;198:812–825. DOI: 10.1111/bjh.18276.
14. Kahn SR, Lamping DL, Ducruet T, Arseneault L, Miron MJ, Roussin A, et al. VEINES-QOL/Sym questionnaire was a reliable and valid disease-specific quality of life measure for deep venous thrombosis. *Journal of Clinical Epidemiology*. 2006;59:1049–1056. DOI: 10.1016/j.jclinepi.2005.10.016.
15. Cohn DM, Nelis EA, Busweiler LA, Kaptein AA, Middeldorp S. Quality of life after pulmonary embolism: the development of the PEmb-QoL questionnaire. *Journal of Thrombosis and Haemostasis*. 2009;7:1044–1046. DOI: 10.1111/j.1538-7836.2009.03341.x.
16. Klok FA, Cohn DM, Middeldorp S, Scharloo M, Büller HR, van Kralingen KW, et al. Quality of life after pulmonary embolism: validation of the PEmb-QoL questionnaire. *Journal of Thrombosis and Haemostasis*. 2010;8:523–532. DOI: 10.1111/j.1538-7836.2009.03726.x.
17. Marin-Barrera L, Muñoz-Martin AJ, Rios-Herranz E, Garcia-Escobar I, Beato C, Font C, et al. A case-control analysis of the impact of venous thromboembolic disease on quality of life of patients with cancer: Quality of life in Cancer (Qca) study. *Cancers (Basel)*. 2020;12:75. DOI: 10.3390/cancers12010075.
18. Farge D, Cajfinger F, Falvo N, Berremili T, Couturaud F, Bensaoula O, et al. Quality of life in cancer patients undergoing anticoagulant treatment with LMWH for venous thromboembolism: the QUAVITEC study. *Oncotarget*. 2018;9:26990–26999. DOI: 10.18632/oncotarget.25454.
19. Benelhaj NE, Hutchinson A, Maraveyas A, Johnson MJ. Cancer patients' experiences of the diagnosis and treatment of incidental pulmonary embolism (a qualitative study). *PLOS ONE*. 2022;17:e0276754. DOI: 10.1371/journal.pone.0276754.
20. Baddeley E, Torrens-Burton A, Newman A, Nelson A, Pease N, Nelson R, et al. A mixed-methods study to evaluate a patient-designed tool to reduce harm from cancer-associated thrombosis: The EMPOWER study. *Research and Practice in Thrombosis and Haemostasis*. 2021;5:e12545. DOI: 10.1002/rth2.12545.
21. Planquette B, Bertolotti L, Charles-Nelson A, Laporte S, Grange C, Mahé I, et al; CASTA DIVA Trial Investigators. Rivaroxaban vs Dalteparin in Cancer-Associated Thromboembolism: A Randomized Trial. *Chest*. 2022;161(3):781–790. DOI: 10.1016/j.chest.2021.09.037.
22. Schrag D, et al. Direct Oral Anticoagulants vs Low-Molecular-Weight Heparin for Treatment of Cancer-Associated Venous Thromboembolism: The CANVAS Randomized Clinical Trial. *JAMA*. 2023. DOI: 10.1001/jama.2023.7843.
23. Potere N, Barco S, Mahé I, et al. Awareness of venous thromboembolism among patients with cancer: Preliminary findings from a global initiative for World Thrombosis Day. *Journal of Thrombosis and Haemostasis*. 2022;20(12):2964–2971. DOI: 10.1111/jth.15902.
24. Mazzolai L, Alatri A. Treatment of cancer-associated venous thromboembolism. *European Society of Cardiology — CardioPractice (educational review)*. 2022. URL: <https://www.escardio.org/CardioPractice> (дата обращения: 14.04.2026).
25. Lutsey PL, Zakai NA. Epidemiology and prevention of venous thromboembolism. *Nature Reviews Cardiology*. 2023;20:248–262. DOI: 10.1038/s41569-022-00787-6.

© Матвеев Дмитрий Валентинович (dvmatveev@mail.ru); Кузнецов Максим Робертович (mrkuznetsov@mail.ru);
 Долидзе Давид Джонович (ddolidzed@mail.ru); Лебедев Сергей Сергеевич (lebedevssd@yandex.ru);
 Чебан Сергей Валерьевич (tchebansergiy@yandex.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

ОСОБЕННОСТИ ВЕДЕНИЯ ПАЦИЕНТОК С ГЕСТАЦИОННЫМ САХАРНЫМ ДИАБЕТОМ И СОПУТСТВУЮЩИМ ОЖИРЕНИЕМ

FEATURES OF MANAGEMENT OF PATIENTS WITH GESTATIONAL DIABETES MELLITUS AND CONCOMITANT OBESITY

A. Sizov

Summary. The combination of gestational diabetes and obesity is one of the most difficult clinical situations in modern obstetrics, since both conditions have a negative impact on the course of pregnancy and outcomes for the mother and fetus. The presented article analyzes the features of managing pregnant women with gestational diabetes mellitus against the background of obesity. The results showed that patients with obesity and gestational diabetes mellitus have more pronounced insulin resistance, higher levels of fasting glycemia, the need for insulin therapy, as well as an increased risk of preeclampsia, macrosomia, and neonatal complications. Special attention is paid to non-drug treatment methods, including diet therapy and physical activity, which are priorities at the initial stage of management. As an author's development, a structured algorithm for the management of pregnant patients with gestational diabetes mellitus and concomitant obesity is proposed. The presented algorithm can be used in clinical practice. The main control points are: timely screening in the first trimester, strict monitoring of gestational weight gain, the preferred use of metformin if necessary for pharmacotherapy (in the absence of contraindications), as well as planned delivery at week 39 with mandatory intranatal glycemic control and prevention of neonatal hypoglycemia. The conclusion is made about the need for a personal approach with an emphasis on lifestyle correction and a rational choice of medicines.

Keywords: gestational diabetes mellitus, GDM, obesity, pregnancy, insulin resistance, diet therapy.

Сизов Алексей Викторович

кандидат медицинских наук, доцент,
Санкт-Петербургский государственный
педиатрический медицинский университет»
Министерства здравоохранения Российской Федерации
avsizov@inbox.ru

Аннотация. Сочетание гестационного сахарного диабета и ожирения представляет собой одну из наиболее сложных клинических ситуаций в современном акушерстве, поскольку оба состояния оказывают негативное влияние на течение беременности и исходы для матери и плода. В представленной статье проведён анализ особенностей ведения беременных с гестационным сахарным диабетом на фоне ожирения. Результаты показали, что у пациенток с ожирением и гестационным сахарным диабетом отмечаются более выраженная инсулинорезистентность, более высокие показатели гликемии натощак, потребность в инсулинотерапии, а также повышенный риск преэклампсии, макросомии плода и неонатальных осложнений. Особое внимание уделено немедикаментозным методам лечения, включая диетотерапию и физическую активность, которые являются приоритетными на начальном этапе ведения. В качестве авторской разработки предложен структурированный алгоритм ведения беременных пациенток с гестационным сахарным диабетом и сопутствующим ожирением. Представленный алгоритм может быть использован в клинической практике. Основными точками контроля являются: своевременный скрининг в первом триместре, строгий контроль гестационной прибавки массы тела, предпочтительное использование метформина при необходимости фармакотерапии (при отсутствии противопоказаний), а также плановое родоразрешение на 39-й неделе с обязательным интранатальным контролем гликемии и профилактикой неонатальной гипогликемии. Сделан вывод о необходимости персонального подхода с акцентом на коррекцию образа жизни и рациональный выбор лекарственных средств.

Ключевые слова: гестационный сахарный диабет, ГСД, ожирение, беременность, инсулинорезистентность, диетотерапия.

Введение

Ожирение и гестационный сахарный диабет (ГСД) представляют собой две наиболее распространённые метаболические патологии периода гестации, причём в последние десятилетия наблюдается их рост во всём мире. По данным Всемирной организации здравоохранения, распространённость ожирения среди женщин репродуктивного возраста достигает 15–20 % в европейских странах и значительно выше в США и странах Ближнего Востока. Одновременно с этим частота ГСД, по данным Международной диабетической федерации, достигает 14 % в мире и до 25 % в отдельных

популяциях [9]. Два указанных состояния часто встречаются одновременно, поскольку ожирение служит одним из самых значимых факторов риска развития ГСД.

Целью настоящей статьи является систематический анализ особенностей ведения беременных с гестационным сахарным диабетом и сопутствующим ожирением, а также разработка практических рекомендаций по оптимизации тактики ведения таких пациенток. Новизна исследования заключается в предложенном алгоритме ведения беременных пациенток с гестационным сахарным диабетом и сопутствующим ожирением. Основными точками контроля являются: своевременный скрининг

в первом триместре, строгий контроль гестационной прибавки массы тела, предпочтительное использование метформина при необходимости фармакотерапии (при отсутствии противопоказаний), а также плановое родоразрешение на 39-й неделе с обязательным интранатальным контролем гликемии и профилактикой неонатальной гипогликемии.

Материалы и методы исследования

Материалами исследования послужили данные различных российских и зарубежных исследований с участием беременных женщин, а также систематических обзоров, клинических рекомендаций и научных статей российских и зарубежных авторов. Методы исследования: сравнение, описание, классификация, сопоставление, контент-анализ, табличный метод.

Результаты и обсуждения

Ожирение и ГСД имеют общие патогенетические механизмы, что объясняет их частое сочетание и взаимное отягощение. В основе обоих состояний лежит инсулинорезистентность — снижение чувствительности периферических тканей к действию инсулина. У женщин с ожирением инсулинорезистентность присутствует уже до наступления беременности, тогда как у женщин с нормальной массой тела она развивается только во втором-третьем триместрах как физиологический механизм, обеспечивающий увеличение поступления глюкозы к плоду. Когда эти два процесса накладываются друг на друга, компенсаторные возможности β -клеток поджелудочной железы оказываются недостаточными, что приводит к ГСД.

Это подтверждается исследованием Catalano, P.M., Kirwan J.P., Haugel-de Mouzon S., King J., которые выявили, что у женщин с ожирением исходно снижена чувствительность к инсулину, а во время беременности, когда инсулинорезистентность закономерно возрастает (примерно на 60 % к третьему триместру), эти изменения становятся ещё более выраженными [6]. У пациенток с ожирением и ГСД отмечаются более высокие показатели гликемии натощак, более значительная гипергликемия после приёма пищи, а также более высокая потребность в инсулинотерапии по сравнению с женщинами с нормальной массой тела [5]. Кроме того, сочетание ожирения и ГСД существенно повышает риск развития преэклампсии, макросомии плода, дистонии плечиков, неонатальной гипогликемии и долгосрочных метаболических нарушений у потомства.

Согласно рекомендациям ВОЗ, основой лечения ГСД, в том числе у пациенток с ожирением, является немедикаментозная терапия — индивидуализированное диетическое консультирование и физическая активность

[9]. При этом подчёркивается, что диетические рекомендации должны следовать общим принципам здорового питания, а не представлять собой специализированную диету для контроля гликемии при беременности. Основные рекомендации включают приоритет цельных злаков, овощей, фруктов и бобовых в качестве основных источников углеводов, потребление не менее 400 г овощей в день, обеспечение суточного потребления пищевых волокон не менее 25 г, а также ограничение свободных сахаров до менее 5–10 % от общего потребления энергии [9].

Особого внимания заслуживает вопрос о допустимой прибавке массы тела во время беременности у женщин с ожирением. Согласно руководству MSD Manual, для женщин с ожирением рекомендуется ограничение гестационной прибавки массы тела до 5–9,1 кг (11–20 фунтов) [8]. При этом, как подчёркивают эксперты ВОЗ, нельзя искусственно подавлять нормальную прибавку веса — цель состоит в достижении оптимальной прибавки, позволяющей снизить риски как макросомии, так и низкой массы тела новорождённого при рождении [9].

В отношении конкретных типов диет анализ L. Zhang показал, что DASH-диета (Dietary Approaches to Stop Hypertension) и низкогликемическая диета наиболее эффективны для контроля гликемии при ГСД. DASH-диета продемонстрировала преимущества в снижении уровня глюкозы натощак (средняя разница — 11,52 мг/дл), а также снижала риск макросомии на 89 % и частоту кесарева сечения на 46 %. Низкогликемическая диета показала преимущества в контроле постпрандиальной гликемии (средняя разница — 6,3 мг/дл). Обе диеты также способствовали снижению массы тела новорождённых [10].

Всем беременным без противопоказаний рекомендуется не менее 150 минут аэробной нагрузки умеренной интенсивности в неделю, а также умеренные силовые тренировки при возможности [9]. Физическая активность способствует улучшению утилизации глюкозы периферическими тканями, снижению инсулинорезистентности и профилактике избыточной гестационной прибавки массы тела. Качество изменения состава тела (особенно снижение висцерального жира) может быть более важным фактором метаболических улучшений, чем просто снижение общей массы тела [7].

Таким образом, ведение беременных с сочетанной патологией — ГСД и ожирением — требует учёта множества факторов, начиная от патогенетических особенностей и заканчивая выбором метода родоразрешения. Особенности ведения данной категории пациенток систематизированы в таблице 1.

Как следует из представленных данных, главным отличием ведения пациенток с ожирением является необ-

Таблица 1.

Особенности ведения пациенток с ГСД и сопутствующим ожирением

Аспект ведения	Особенности у пациенток с ожирением	Рекомендации
Патогенез	Более выраженная исходная инсулинорезистентность, более высокий уровень гликемии натощак, преобладание изолированного нарушения гликемии натощак	Ранний скрининг ГСД (предпочтительно в I триместре), особое внимание контролю гликемии натощак
Гестационная прибавка массы тела	Повышенный риск избыточной прибавки, что усугубляет метаболические нарушения	Рекомендуемая прибавка 5–9,1 кг; ежемесячный контроль массы тела; структурированные программы питания
Диетотерапия	Требуется более строгий контроль калорийности и качества углеводов; риск нутритивной недостаточности при чрезмерных ограничениях	DASH-диета или низкогликемическая диета; не менее 25 г пищевых волокон в сутки; ограничение свободных сахаров <5–10 % калорийности
Физическая активность	Снижение физической активности из-за большей массы тела и сопутствующих жалоб (боли в спине, одышка)	Не менее 150 минут умеренной аэробной нагрузки в неделю; силовые упражнения
Выбор фармакотерапии	Более высокая потребность в инсулине; метформин связан с меньшей прибавкой массы тела	Метформин как предпочтительный вариант при отсутствии противопоказаний; инсулин при тяжёлой гипергликемии
Мониторинг гликемии	Более высокий риск гипергликемии натощак, требующий более частого контроля	Самоконтроль гликемии ежедневно; при нестабильных показателях — до 4 раз в сутки (натощак и после каждого приёма пищи)
Фетальный мониторинг	Более высокий риск макросомии, преэклампсии и мертворождения	УЗИ роста плода каждые 4 недели после 28 недель; доплерометрия; нестрессовый тест с 32 недель
Срок родоразрешения	Более высокий риск осложнений при пролонгировании беременности	При медикаментозно контролируемом ГСД — 39 0/7–39 6/7 недель; при макросомии >4500 г — кесарево сечение

Источник: составлено автором на основе данных [1–4, 6, 8–10]

ходимость более раннего скрининга, особый контроль гликемии натощак (которая у них повышена в большей степени, чем постприандиальная гликемия), а также предпочтительное использование метформина как препарата, связанного с меньшей прибавкой массы тела. Эти особенности должны учитываться при разработке индивидуального плана ведения каждой пациентки.

Когда немедикаментозные методы не позволяют достичь целевых показателей гликемии (глюкоза натощак >95 мг/дл или постприандиальная гликемия >120 мг/дл в течение 1–2 недель диетотерапии), требуется назначение фармакотерапии. Главным вопросом для пациенток с ожирением является выбор между метформином и инсулином. Согласно рекомендациям ВОЗ, как метформин, так и инсулин могут рассматриваться в качестве препаратов первого ряда для лечения ГСД [9]. Однако существуют исследования, показывающие, что пероральные препараты ассоциированы с меньшей прибавкой веса у пациенток с ГСД по сравнению с инсулином [4]. Согласно руководству MSD Manual, для женщин с ожирением рекомендуется приём короткодействующего инсулина перед каждым приёмом пищи [8].

Кроме того, всем беременным с ГСД рекомендуется самоконтроль гликемии с частотой, определяемой индивидуально в зависимости от типа диабета, приме-

нения фармакотерапии и стабильности гликемического контроля. Что касается непрерывного мониторинга глюкозы, ВОЗ не рекомендует его рутинное применение при ГСД ввиду ограниченных доказательств улучшения перинатальных исходов, однако допускает его использование при наличии ресурсов и потенциальной пользы для конкретной пациентки. Также следует отметить, что у пациенток с ГСД и ожирением отмечается более высокий риск макросомии плода, что требует тщательного ультразвукового мониторинга. Согласно рекомендациям ВОЗ, все беременные с диабетом должны пройти обязательное детальное ультразвуковое исследование до 24 недель гестации [9].

На основе проанализированных данных и с учётом выявленных особенностей ведения пациенток с ожирением нами был разработан структурированный алгоритм ведения беременных с ГСД и ожирением. Алгоритм охватывает все этапы — от скрининга в первом триместре до послеродового наблюдения, интегрируя немедикаментозные и медикаментозные методы лечения, а также фетальный мониторинг и выбор оптимального срока родоразрешения. Предлагаемый алгоритм представлен в таблице 2.

Представленный алгоритм может быть использован в клинической практике для стандартизации ведения

Таблица 2.

Предлагаемый алгоритм ведения беременных с ГСД и ожирением

Этап	Срок гестации	Мероприятия
I этап: Скрининг и диагностика	I триместр (8–14 недель)	Определение гликемии натощак у всех беременных с ожирением ($IMT \geq 30 \text{ кг/м}^2$); при выявлении нарушений — пероральный глюкозотолерантный тест с 75 г глюкозы; оценка исходного индекса массы тела, артериального давления, липидного профиля
II этап: Немедикаментозная терапия	Весь период беременности	Индивидуализированное диетическое консультирование (DASH-диета или низкогликемическая диета с акцентом на контроль гликемии натощак); физическая активность (не менее 150 мин/нед, сочетание аэробных и силовых нагрузок); контроль гестационной прибавки массы тела (целевая прибавка 5–9,1 кг); самоконтроль гликемии (натощак и через 1 час после каждого приёма пищи)
III этап: Медикаментозная терапия	При недостижении целевой гликемии через 1–2 недели диетотерапии	Стартовая терапия: метформин 500 мг 1–2 раза в сутки, при необходимости титрация до 2000–2500 мг/сут; при неэффективности или непереносимости метформина — добавление инсулина (короткодействующий перед едой, особенно при постпрандиальной гипергликемии; пролонгированный инсулин на ночь при гипергликемии натощак); при тяжёлой гипергликемии (глюкоза натощак >110 – 120 мг/дл) — инсулин в качестве терапии первой линии
IV этап: Фетальный мониторинг	С 24–28 недель до родоразрешения	УЗИ плода с оценкой фетометрии и предполагаемой массы тела каждые 4 недели; доплерометрия маточно-плацентарного и плодово-плацентарного кровотока; нестрессовый тест или биофизический профиль плода с 32 недель (еженедельно)
V этап: Родоразрешение	39 0/7–39 6/7 недель при медикаментозно контролируемом ГСД; 39–40 6/7 недель при ГСД класса A1	Плановое родоразрешение в указанные сроки; при макросомии плода (предполагаемая масса $>4500 \text{ г}$) — плановое кесарево сечение; интранатальный контроль гликемии (каждый час); профилактика неонатальной гипогликемии (раннее прикладывание к груди, контроль гликемии новорождённого через 2, 4 и 6 часов после рождения)
VI этап: Послеродовое наблюдение	6–12 недель после родов	Пероральный глюкозотолерантный тест с 75 г глюкозы для исключения манифестного СД 2 типа или нарушенной толерантности; рекомендации по грудному вскармливанию (снижает риск СД 2 типа на 15–30 %); программа снижения массы тела для женщин с ожирением (цель — снижение на 7–10 % от исходной массы тела)

Источник: составлено автором.

беременных с ГСД и ожирением. Основными точками контроля являются: своевременный скрининг в первом триместре, строгий контроль гестационной прибавки массы тела, предпочтительное использование метформина при необходимости фармакотерапии (при отсутствии противопоказаний), а также плановое родоразрешение на 39-й неделе с обязательным интранатальным контролем гликемии и профилактикой неонатальной гипогликемии.

Выводы

Проведённый анализ позволяет сформулировать следующие выводы. Сочетание гестационного сахарного диабета и ожирения характеризуется более выраженной инсулинорезистентностью, более высоким уровнем гликемии натощак, большей потребностью в инсулинотерапии и повышенным риском материнских и перинатальных осложнений. Патогенетически эти два состо-

яния тесно взаимосвязаны, что требует комплексного подхода к ведению.

В статье проведён анализ особенностей ведения беременных с гестационным сахарным диабетом на фоне ожирения. Результаты показали, что у пациенток с ожирением и гестационным сахарным диабетом отмечают более выраженная инсулинорезистентность, более высокие показатели гликемии натощак, потребность в инсулинотерапии, а также повышенный риск преэклампсии, макросомии плода и неонатальных осложнений. Особое внимание уделено немедикаментозным методам лечения, включая диетотерапию и физическую активность, которые являются приоритетными на начальном этапе ведения. В качестве авторской разработки предложен алгоритм ведения беременных пациенток с гестационным сахарным диабетом и сопутствующим ожирением. Представленный алгоритм может быть использован в клинической практике для стандартизации

ведения беременных с ГСД и ожирением. Основными точками контроля являются: своевременный скрининг в первом триместре, строгий контроль гестационной прибавки массы тела, предпочтительное использование метформина при необходимости фармакотерапии (при

отсутствии противопоказаний), а также плановое родоразрешение на 39-й неделе с обязательным интранатальным контролем гликемии и профилактикой неонатальной гипогликемии.

ЛИТЕРАТУРА

1. Блох, Р.А., Блох М.Е., Мухамедрахимов Р.Ж. Психологические и клинические особенности беременности, родов и послеродового периода у женщин с гестационным сахарным диабетом // Национальный психологический журнал. — 2026. — Т. 21. — № 1. — С. 19–29. — DOI: 10.11621/npj.2026.0102.
2. Османова, С.Д., Цахилова С.Г., Царькова М.А., Сащенко А.А. Ожирение как фактор развития гестационного сахарного диабета // Фарматека. — 2022. — № 6. — С. 76–79. <https://dx.doi.org/10.18565/pharmateca.2022.6.76-79>.
3. Пирматова, Д.А., Додхоева М.Ф., Пархофер К.Г. Особенности течения беременности и родов у женщин с гестационным сахарным диабетом // Медицинский вестник Национальной академии наук Таджикистана. — 2020. — Т. 10, № 2 (34). — С.163–167.
4. Amaral-Moreira, C.F.A., Paulino D.S.M., Guida J.P.S., Pereira B.G., Rehder P.M., Surita F.G. Gestational weight gain according to treatment in gestational diabetes: a systematic review and meta-analysis // Revista Brasileira de Ginecologia e Obstetrícia. — 2025. — Vol. 47. — pp. 65–72.
5. American Diabetes Association Professional Practice Committee. 15. Management of Diabetes in Pregnancy: Standards of Care in Diabetes–2026 // Diabetes Care. — 2026. — Vol. 49. — Suppl. 1. — pp. S321–S338. — DOI: 10.2337/dc26-S015.
6. Catalano, P.M., Kirwan J.P., Haugel-de Mouzon S., King J. Gestational diabetes and insulin resistance: role in short— and long-term implications for mother and fetus // The Journal of Nutrition. — 2003. — Vol. 133. — № 5. — pp. 1674S–1683S. — DOI: 10.1093/jn/133.5.1674S.
7. Fujikawa T. Reconsidering the Quality of Weight Change in Gestational Diabetes Prevention // BMJ. — 2025. — №390. — pp. e083398.
8. Management of Gestational Diabetes During Pregnancy [Electronic resource] // MSD Manual Professional Edition. — 2026. — URL: <https://www.msdmanuals.com/professional/multimedia/table/management-of-gestational-diabetes-during-pregnancy> (дата обращения: 10.04.2026).
9. WHO recommendations on care for women with diabetes during pregnancy. — Geneva: World Health Organization, 2025. — 89 p.
10. Zhang, L., Wang F., Tashiro S., Liu P.J. Effects of Dietary Approaches and Exercise Interventions on Gestational Diabetes Mellitus: A Systematic Review and Bayesian Network Meta-analysis // American Journal of Clinical Nutrition. — 2024. — Vol. 120. — № 6. — pp. 1245–1258. — DOI: 10.1016/j.ajcnut.2024.08.015.

© Сизов Алексей Викторович (avsizov@inbox.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

ЗАВИСИМОСТЬ ЧАСТОТЫ ПРЕЖДЕВРЕМЕННЫХ И ОПЕРАТИВНЫХ РОДОВ ОТ НАЛИЧИЯ ЭКСТРАГЕНИТАЛЬНОЙ ПАТОЛОГИИ У ПАЦИЕНТОК С ПРЕЭКЛАМПСИЕЙ РАЗЛИЧНОЙ СТЕПЕНИ ТЯЖЕСТИ

DEPENDENCE OF THE FREQUENCY OF PREMATURE AND OPERATIVE LABOR ON THE PRESENCE OF EXTRAGENITAL PATHOLOGY IN PATIENTS WITH PREECLAMPSIA OF VARIOUS DEGREES OF TENSION

I. Simanov

Summary. The frequency of preterm birth in «moderate» preeclampsia was largely determined by the presence of extragenital pathology. In the groups of patients with «moderate» preeclampsia against the background of extragenital diseases, premature birth occurred 2 times more often than without diseases. The number of operative births in the study group depended on the severity of preeclampsia. In «severe» preeclampsia, operative labor was observed 2–3 times more often than in «moderate» preeclampsia. The maximum number of operations was performed in «severe» preeclampsia against the background of extragenital diseases. In the group of healthy patients, preterm birth was observed exclusively in persons with a high infectious index, and the frequency of cesarean section operations did not differ from the rates in the general population. The main indications for surgery in healthy patients were complications in childbirth.

Keywords: cesarean section, preeclampsia, premature birth.

Симанов Игорь Викторович

кандидат медицинских наук, доцент,
Пятигорский медико-фармацевтический институт —
филиал ФГБОУ ВО ВолгГМУ Минздрава России
simigor1@rambler.ru

Аннотация. Частота преждевременных родов при «умеренной» преэклампсии в большей степени определялась наличием экстрагенитальной патологии. В группах пациенток с «умеренной» преэклампсией на фоне экстрагенитальных заболеваний преждевременные роды встречались в 2 раза чаще, чем без заболеваний. Количество оперативных родов в основной группе зависело от степени тяжести преэклампсии. При «тяжелой» преэклампсии в 2–3 раза чаще, чем при «умеренной» наблюдались оперативные роды. Максимальное количество операций было выполнено при «тяжелой» преэклампсии на фоне экстрагенитальных заболеваний.

В группе здоровых пациенток преждевременные роды наблюдались исключительно у лиц с высоким инфекционным индексом, а частота операций кесарева сечения не отличалась от показателей в общей популяции. Основными показаниями к операции у здоровых пациенток были осложнения в родах.

Ключевые слова: кесарево сечение, преэклампсия, преждевременные роды.

Введение

Актуальность данного исследования не вызывает сомнений поскольку за последнее десятилетие увеличивается количество преждевременного и оперативного родоразрешения при преэклампсии. Несмотря на полученные многочисленные данные о частоте преждевременных родов при преэклампсии, недостаточно изучена причина этого повышения. Ряд ученых считают, что на развитие данного процесса при преэклампсии играет роль степень ее тяжести, другие — придерживаются мнения о роли тяжелой экстрагенитальной патологии, третьи — вообще не видят связь между преэклампсией и преждевременными родами [2, с. 12; 3, с. 68].

Задачами данного исследования не являлось получение каких-либо новых открытий по частоте преждевре-

менных родов при преэклампсии, поскольку эти данные давно всем известны. Несмотря на полученные многочисленные данные по этому вопросу, преждевременные роды за последнее десятилетие занимают лидирующее положение в группах пациенток с преэклампсией.

Цель исследования

Цель исследования — изучить зависимость частоты преждевременных и оперативных родов от наличия экстрагенитальной патологии при преэклампсии.

Материалы и методы исследования

Количество обследованных представлено в табл. 1.

Из основной группы исследования исключены пациентки, с острыми, подострыми и хроническими инфек-

Таблица 1.
Количество обследованных пациенток с преэклампсией

Степень тяжести преэклампсии	Анамнез не отягощен (n=80) n (%)	Отягощенный анамнез (n=220) n (%)
«умеренная» (n=237)	80(100 %)	157 (71,4 %)
«тяжелая» (n=63)	—	63(28,6 %)

циями различного генеза, имеющие патологию шейки матки и пороки развития матки, многоплодную и многоводную беременность, поскольку эти факторы являются причиной преждевременных родов.

Результаты исследования и их обсуждение

Большое диагностическое значение имеет возраст пациенток с преэклампсией, поскольку у «юных» недостаточно развита лимбическая система, а у «возрастных» присоединяется большое количество экстрагенитальных заболеваний.

Возрастной состав обследуемых пациенток был достаточно разнородным (табл. 2).

Таблица 2.
Возраст исследуемых пациенток

Возраст	«умеренная» преэклампсия n (%)	«тяжелая» n (%)	Контрольная группа n (%)	(p)
1. до 18 лет (n=30)	20 (8,4 %)	10 (16,7 %)	—	—
2. от 19 до 30 лет (n=402)	184 (77,6 %)	33 (55 %)	185 (92,5 %)	0,03
3. старше 30 лет (n=68)	33 (14 %)	20 (31,8 %)	15 (7,5 %)	0,01
Итого: (n=500)	237	63	200	—

Данные, представленные в табл. 2 свидетельствуют о том, что при «умеренной» преэклампсии возрастной состав большинством (77,6 %) представлен пациентками фертильного возраста.

При «тяжелой» преэклампсии имели место быть «юные» и «возрастные» первородящие в количестве 16,7 % и 31,8 %, соответственно.

У пациенток контрольной группы «юных» не было, а «возрастные» наблюдались в незначительном количестве (7,5 %).

Это еще раз доказывает тот факт, что преэклампсия чаще всего развивается у лиц фертильного возраста. «Юный» и возраст старше 30 лет является неким барьером развития тяжелых форм данной патологии [5, с. 85; 8, с. 70; 10, с. 15].

На современном этапе следует отметить рост числа хронических заболеваний у «возрастных» пациенток с преэклампсией. Хронические соматические заболевания являются патогенетическими предикторами развития тяжелых форм преэклампсии, а нередко приводят к приступам эклампсии [6, с. 14; 9, с. 10].

Полученные данные проведенного исследования убедительно показали большой процент заболеваний у данных лиц (табл. 3).

Таблица 3.
Соматическая патология у пациенток исследуемой группы

Заболевания	«умеренная» преэклампсия n (%)	«тяжелая» n (%)	Контрольная группа n (%)	(p)
1. Нарушение жирового обмена (n=118)	64 (40,8 %)	14 (22,2 %)	40 (20 %)	0,02
2. Пиелонефрит (n=85)	55 (35 %)	30 (47,6 %)	—	—
3. Гипертоническая болезнь (n=57)	38 (24,2 %)	19 (30,2 %)	—	—
4. Без патологии (n=160)	—	—	160 (80 %)	—
Итого: (n=420)	157	63	200	—

Данные табл. 3 показали, что у пациенток с преэклампсией в анамнезе преобладали заболевания сосудисто-почечного и эндокринного генеза.

Практически каждая третья пациентка с преэклампсией и каждая пятая — с неосложненной беременностью имела клинические проявления нарушения жирового обмена различной степени выраженности.

Таким образом, такие заболевания, как гипертоническая болезнь, заболевания почек играют важную роль в патогенезе преэклампсии и дородовое лечение данной патологии снизит количество преждевременного и оперативного родоразрешения.

Большую актуальность для современного акушерства представляет собой преждевременные роды, поскольку они представляют опасность для плода и являются

предиктором развития младенческой заболеваемости и смертности [7, с. 70; 11, с. 114; 13, с. 290].

В проведенном исследовании количество преждевременных родов у пациенток с преэклампсией зависело от наличия экстрагенитальных заболеваний.

При гипертонической болезни у пациенток с «умеренной» преэклампсией ранние преждевременные роды были выявлены у каждой десятой пациентки, при заболеваниях почек — у каждой пятой, при нарушении жирового обмена — у каждой третьей.

Ранние преждевременные роды были у каждой третьей пациентки при «тяжелой» преэклампсии на фоне гипертонической болезни, у 5.7 % — очень ранние преждевременные роды на фоне заболеваний почек.

Очень ранние преждевременные роды также наблюдались у каждой десятой пациентки при «тяжелой» преэклампсии на фоне гипертонической болезни, у 12.3 % — при нарушении жирового обмена.

Ранние преждевременные роды у половины обследуемых с «умеренной» преэклампсией протекали с осложнениями, у другой половины — без осложнений.

Среди осложнений родов у каждой пятой из них преобладала первичная слабость родовой деятельности, у каждой десятой — несвоевременное излитие околоплодных вод.

Это доказывает тот факт, что преждевременные роды являются фактором риска развития аномалий родовой деятельности.

На современном этапе увеличивается количество пациенток с оперативным родоразрешением [1, с. 448; 4, с. 40; 12, с. 15].

Проведенное исследование показало, что количество операций кесарево сечение при преэклампсии зависело от ее степени тяжести и наличия экстрагенитальной патологии.

Так, анализ данных показал, что при преэклампсии на фоне гипертонической болезни в 2–3 раза чаще от-

мечались оперативные роды.

При заболеваниях почек и нарушениях жирового обмена оперативное родоразрешение было у каждой третьей пациентки.

Причем, максимальное количество операций было выполнено при «тяжелой» преэклампсии. У 25 % из них «тяжелая» степень данной патологии сочеталась с декомпенсацией плацентарной недостаточности, у 15 % — с преждевременной отслойкой нормально расположенной плаценты, у 23 % — с рубцом на матке после операции кесарево сечение.

У большинства обследуемых при «умеренной» преэклампсии показатели оперативного родоразрешения не отличались от общепопуляционных значений.

У пациенток с «тяжелой» преэклампсией ситуация другая, здесь главнейшую роль играла тяжелая экстрагенитальная патология, которая осложняла течение беременности и приводила не только к ухудшению общего состояния беременной, но и плода.

У большинства пациенток контроля показатели преждевременного и оперативного родоразрешения были в пределах значений в общей популяции. Ранние преждевременные роды наблюдались у 12.5 % пациенток с высоким инфекционным индексом.

Заключение

Экстрагенитальная патология при преэклампсии отягощает течение родов, приводит к увеличению количества преждевременного и оперативного родоразрешения. Так, анализ данных показал, что при преэклампсии на фоне гипертонической болезни в 2–3 раза чаще отмечались оперативные роды. При заболеваниях почек и нарушениях жирового обмена оперативное родоразрешение было у каждой третьей пациентки.

Причем, максимальное количество операций было выполнено при «тяжелой» преэклампсии.

Знание данных процессов поможет улучшить перинатальные показатели.

ЛИТЕРАТУРА

1. Алиева А.Я., Баладжанова А.Б. Структура показаний к операции кесарево сечение //Бюллетень медицинских интернет-конференций. — 2014. — Т. 4. — № 5. — С. 448.
2. Барановская Е.И., Лосицкая О.А., Войтехович Ю.А. Преэклампсия и преждевременные роды //Современные перинатальные медицинские технологии в решении проблем демографической безопасности. — 2017. — № 10. — С. 11–14.
3. Девятова Е.А. Вечные проблемы акушерства — преэклампсия, преждевременные роды //Акушерство и гинекология. Новости. Мнения. Обучение. — 2014. — № 3 (5). — С. 66–70.

4. Исаева Э.В., Рыскельдиева В.Т. Кесарево сечение при тяжелой преэклампсии: особенности ранней неонатальной адаптации //Российский вестник перинатологии и педиатрии. — 2021. — Т. 66. — № 4. — С. 39–44.
5. Космуратова Ш.Б., Битемирова Ш.К., Жакиева Ш.С., Жылкайдар Г.М., Кайсажанова Г.А. Анамnestические факторы развития преэклампсии //Reproductive Medicine. — 2024. — №2. — С. 80–87.
6. Мухсинов Ф.Б., Худоярова Д.Р., Кобилова З.А.К. Течение беременности с преэклампсией на фоне экстрагенитальных заболеваний //Вестник науки и образования. — 2020. — № 14-2 (92). — С. 57–60.
7. Саиджалилова Д.Д., Эрматова Х.А.К. Некоторые аспекты развития преждевременных родов //Биология и интегративная медицина. — 2016. — № 5. — С. 69–73.
8. Слободина А.В., Рудакова Е.Б., Долгих В.Т., Толкач А.Б., Нейман Т.П. Факторы риска развития тяжелой преэклампсии //Сибирский медицинский журнал (Иркутск). — 2012. — Т.112. — № 5. — С. 68–71.
9. Соколовская Т.А., Ступак В.С. Заболеваемость беременных женщин в Российской Федерации: тенденции и прогнозы //Российский вестник акушера-гинеколога. — 2022. — Т. 22. — № 5. — С. 7–14.
10. Соловьева А.В., Спицына М.А., Доронина О.К., Алейникова Е.Ю. Тяжелая преэклампсия. Клинико-анамnestическая характеристика женщин с индуцированными преждевременными родами //Акушерство и гинекология. Новости. Мнения. Обучение. — 2024. — Т. 12. — № 3 (45). — С. 13–19.
11. Спицына М.А., Соловьева А.В., Алейникова Е.Ю., Мамчик Д.С. Тяжелая преэклампсия. Прогнозирование индуцированных преждевременных родов //Акушерство и гинекология. Новости. Мнения. Обучение. — 2024. — Т. 12. — № 5. — С. 111–115.
12. Ходжаева Д.Н., Аюпова Д.А., Мадалимова Н.Х.К. Выбор тактики родоразрешения при тяжелой преэклампсии //Re-Health Journal. — 2021. — 2(10). — С. 10–17.
13. Шаджанова Н.Ш., Саркисова Л.В., Сафарова Д.К. Причины преждевременных родов и пути их предупреждения //Новый день в медицине. — 2019. — № 3(27). — С. 289–292.

© Симанов Игорь Викторович (simigor1@rambler.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»

Наши авторы

Adaev N. — Samara National Research University named after Academician S.P. Korolev

Alkattan Sameh — postgraduate student, Federal State Budgetary Educational Institution of Higher Education «Saint Petersburg State University of Telecommunications named after Professor M. A. Bonch-Bruевич»

Apana Kenneth Ayinbuno — Postgraduate student, Saint Petersburg Electrotechnical University «LETI»

Bakhteeva G. — Ph.D., Associate Professor, Saratov State Medical University named after V.I. Razumovsky

Balakina D. — Assistant, Saratov State Medical University named after V.I. Razumovsky

Balter R. — Doctor of Medical Sciences, Professor, Samara State Medical University

Cheban S. — Surgeon, Reutov Clinical Hospital

Chelnokov V. — Doctor of Technical Sciences, Professor, D. Mendeleev University of Chemical Technology of Russia, Moscow

Cherenkova I. — Senior Lecturer, K.I. Skryabin Moscow State Academy of Veterinary Medicine and Biotechnology

Cherepenin V. — Graduate student, South Russian State Polytechnic University (NPI) named after M.I. Platova (Novocherkassk)

Chmilevskaya O. — General Director, ANO DPO «Institute of Organizational Development and Leadership»

Conde Abdoul Karim — Postgraduate student, Moscow State University of Technology «STANKIN»

Daudova R. — Associate Professor, Dagestan State Pedagogical University named after R. Gamzatov (Makhachkala)

Dolidze D. — MD, Academic Secretary, Head of the Scientific and Clinical Department, S.P. Botkin Medical Research and Clinical Center, Moscow Department of Health; Professor, Russian Medical Academy of Postgraduate Education, Ministry of Health of the Russian Federation

Dzhagaryan P. — graduate student, Saratov State Medical University named after V.I. Razumovsky

Egorova A. — Junior Researcher, Research Institute for Problems of Ecology and Mineral Wealth Use of Tatarstan Academy of Sciences, Kazan

Our authors

Erokina N. — Doctor of Medical Sciences, Professor, Saratov State Medical University named after V.I. Razumovsky

Esipov I. — Senior Lecturer, Federal State Budgetary Educational Institution of Higher Education «MIREA, Russian Technological University»

Evdoshenko O. — Ph.D. of Engineering Sciences, assistant professor, MIREA — Russian Technological University (Moscow)

Galushkina D. — Graduate student, South Russian State Polytechnic University (NPI) named after M.I. Platova (Novocherkassk)

Gatiyatullina A. — Junior Researcher, Research Institute for Problems of Ecology and Mineral Wealth Use of Tatarstan Academy of Sciences, Kazan

Gerasimenko A. — junior researcher, Rostov-on-Don Antiplague Scientific Research Institute of Rospotrebnadzor

Glukhov A. — JSC «Scientific and Production Association «Critical Information Systems»

Gorokh A. — junior researcher, Rostov-on-Don Antiplague Scientific Research Institute of Rospotrebnadzor

Goryachkin B. — candidate of technical Sciences, associate Professor, Moscow State Technical University names Bauman

Gritsenko D. — postgraduate student, Bauman Moscow State Technical University

Ilchenko O. — Assistant, Samara State Medical University

Ivanova T. — Candidate of Medical Sciences, Associate Professor, Samara State Medical University

Kalinnikova T. — PhD in Biology, Head of the Laboratory, Research Institute for Problems of Ecology and Mineral Wealth Use of Tatarstan Academy of Sciences, Kazan

Katasonov D. — Senior research fellow, Institute of Automation and Electrometry, Siberian Branch of the Russian Academy of Sciences; Associate Professor, Novosibirsk State Technical University, Novosibirsk,

Kishkinova O. — Senior Lecturer, K.I. Skryabin Moscow State Academy of Veterinary Medicine and Biotechnology

Klementyeva O. — Leading Technical Control Expert, ANO DPO «Institute of Organizational Development and Leadership»

Knyazev I. — Scientific Researcher, Research Institute for Problems of Ecology and Mineral Wealth Use of Tatarstan Academy of Sciences, Kazan

Kochemirov M. — Saint Petersburg State University of Aerospace Instrumentation

Koriagin V. — resident, Perm State Medical University named after Academician E.A. Wagner

Koriagina N. — Doctor of Medical Sciences, Professor, Perm State Medical University named after Academician E.A. Wagner

Kotelnikov V. — Applicant, Samara State Medical University

Kozlov A. — Saint Petersburg State University of Aerospace Instrumentation

Kucherenko S. — Applicant, Perm State Medical University named after Academician E.A. Wagner

Kutlikova I. — Senior Lecturer, K.I. Skryabin Moscow State Academy of Veterinary Medicine and Biotechnology

Kuznetsov M. — MD, Professor, L.L. Levshin Institute of Cluster Oncology, Sechenov University; S.S. Yudin City Clinical Hospital,

Kuznetsova T. — Scientific Researcher, Research Institute for Problems of Ecology and Mineral Wealth Use of Tatarstan Academy of Sciences, Kazan

Laskova Yu. — postgraduate student, K.G. Razumovsky Moscow State University of Technologies and Management (the First Cossack University)

Lebedev S. — MD, Professor, Russian Medical Academy of Postgraduate Education, Ministry of Health of the Russian Federation; Deputy Chief Physician for Oncology, S.P. Botkin Medical Research and Clinical Center, Moscow Department of Health

Magomedova U. — Senior Lecturer, Dagestan State Medical Academy (Makhachkala)

Malakhov S. — PhD, Associate Professor, Samara National Research University named after Academician S.P. Koroleva

Maleev O. — PhD, Associate Professor, Peter the Great St. Petersburg Polytechnic University

Markaryan P. — Samara National Research University named after Academician S.P. Korolev

Markin E. — The Kosygin State University of Russia

Matveev D. — MD, Professor, Russian Medical Academy of Postgraduate Education, Ministry of Health of the Russian Federation

Mironov A. — Ph.D., Saratov State Medical University named after V.I. Razumovsky

Mirsagatov D. — Povolzhsky State University of Telecommunications and Informatics, Samara, Russia

Molkova L. — Federal State Autonomic Educational Institution of Higher Education «Peter the Great St. Petersburg Polytechnic University (SPbPU)»

Moskaleva V. — Novosibirsk State Technical University, Novosibirsk

Ngoula Brintch-Geoffrey — PhD Student, Tyumen Industrial University

Nurmagomedova S. — Associate Professor, Dagestan State Medical Academy (Makhachkala)

Otrubyannikov E. — Candidate of Technical Sciences, Associate Professor The Kosygin State University of Russia

Panov G. — Moscow State Technical University names Bauman

Parfenov A. — Assistant, Saratov State Medical University named after V.I. Razumovsky

Parfentiev K. — Ph.D. in Engineering, associate professor, Bauman Moscow State Technical University

Petrov A. — PhD in Biology, Head of the laboratory, Research Institute for Problems of Ecology and Mineral Wealth Use of Tatarstan Academy of Sciences, Kazan

Petrov D. — Postgraduate, Chuvash State University, Cheboksary

Petrov M. — Chuvash State University, Cheboksary

Pikina N. — Candidate of Pedagogical Sciences, associate professor, Moscow Polytechnic University, Cheboksary Institute (branch), Russia, Cheboksary

Pisanov R. — PhD in Biology, leading researcher, Rostov-on-Don Anti plague Scientific Research Institute of Rospotrebnadzor

Potashin A. — Specialist, RTU MIREA

Pozhar K. — Ph.D. Eng., associate professor, National Research University of Electronic Technology (Moscow)

Pozhidaev N. — Samara National Research University named after Academician S.P. Korolev

Rabotaev G. — Moscow Polytechnic University, Cheboksary Institute (branch), Russia, Cheboksary

Savelyeva S. — Ph.D., Associate Professor, Saratov State Medical University named after V.I. Razumovsky

Shatunova V. — National University of Science and Technology MISIS, Moscow, Russian Federation

Shibichenko M. — postgraduate student, Moscow University of Finance and Law MFUA

Shchikina Yu. — Doctor of Technical Sciences, Professor, Saint Petersburg Electrotechnical University «LETI»

Shmakov V. — PhD in Technical Sciences, Associate Professor, Peter the Great St. Petersburg Polytechnic University

Simanov I. — Candidate of Medical Sciences, Associate Professor, Pyatigorsk Medical and Pharmaceutical Institute — Branch of the Volgograd State Medical University of the Ministry of Health of the Russian Federation

Sizov A. — Candidate of Medical Sciences, Associate Professor, Saint-Petersburg State Pediatric Medical University Ministry of Health of the Russian Federation

Staroverov I. — Senior Lecturer, Federal State Budgetary Educational Institution of Higher Education «MIREA, Russian Technological University»

Suleymanova R. — Associate Professor, Dagestan State Medical Academy (Makhachkala)

Sysa P. — Moscow State Technical University named after N.E. Bauman

Taranets M. — Samara National Research University named after Academician S.P. Korolev

Terekhova M. — The Kosygin State University of Russia

Toichkin N. — candidate of technical sciences, associate professor, Murmansk Arctic university, Apatity

Trunova S. — Associate Professor, Dagestan State Medical Academy (Makhachkala)

Tsai Tsin — Postgraduate student, ITMO University

Tselkovich L. — Doctor of Medical Sciences, Professor, Samara State Medical University

Tchipungu Eduardo Afonso — Postgraduate Student, National University of Science and Technology MISIS, Moscow, Russian Federation

Ushakov D. — Saint Petersburg State University of Aerospace Instrumentation

Uzbekov R. — graduate student, D. Mendeleev University of Chemical Technology of Russia, Moscow

Vasin N. — Peter the Great St. Petersburg Polytechnic University

Vodopyanov A. — PhD in Medicine, leading researcher, Rostov-on-Don Antiplague Scientific Research Institute of Rospotrebnadzor

Vorobyov E. — Moscow Polytechnic University, Cheboksary Institute (branch), Russia, Cheboksary

Vorobyov S. — Assistant professor, South Russian State Polytechnic University (NPI) named after M.I. Platova (Novocherkassk)

Voronin S. — Candidate of Law, Professor, Rector of the National Research Institute

Zakharova N. — Doctor of Medical Sciences, Professor, Saratov State Medical University named after V.I. Razumovsky

Требования к оформлению статей, направляемых для публикации в журнале



Для публикации научных работ в выпусках серий научно-практического журнала «Современная наука: актуальные проблемы теории и практики» принимаются статьи на русском языке. Статья должна соответствовать научным требованиям и общему направлению серии журнала, быть интересной достаточно широкому кругу российской и зарубежной научной общественности.

Материал, предлагаемый для публикации, должен быть оригинальным, не опубликованным ранее в других печатных изданиях, написан в контексте современной научной литературы, и содержать очевидный элемент создания нового знания. Представленные статьи проходят проверку в программе «Антиплагиат».

За точность воспроизведения дат, имен, цитат, формул, цифр несет ответственность автор.

Редакционная коллегия оставляет за собой право на редактирование статей без изменения научного содержания авторского варианта.

Научно-практический журнал «Современная наука: актуальные проблемы теории и практики» проводит независимое (внутреннее) рецензирование.

Правила оформления текста.

- ◆ Текст статьи набирается через 1,5 интервала в текстовом редакторе Word для Windows с расширением ".doc", или ".rtf", шрифт 14 Times New Roman.
- ◆ Перед заглавием статьи указывается шифр согласно универсальной десятичной классификации (УДК).
- ◆ Рисунки и таблицы в статью не вставляются, а даются отдельными файлами.
- ◆ Единицы измерения в статье следует выражать в Международной системе единиц (СИ).
- ◆ Все таблицы в тексте должны иметь названия и сквозную нумерацию. Сокращения слов в таблицах не допускаются.
- ◆ Литературные источники, использованные в статье, должны быть представлены общим списком в ее конце. Ссылки на упомянутую литературу в тексте обязательны и даются в квадратных скобках. Нумерация источников идет в последовательности упоминания в тексте.
- ◆ Литература составляется в соответствии с ГОСТ 7.1-2003.
- ◆ Ссылки на неопубликованные работы не допускаются.

Правила написания математических формул.

- ◆ В статье следует приводить лишь самые главные, итоговые формулы.
- ◆ Математические формулы нужно набирать, точно размещая знаки, цифры, буквы.
- ◆ Все использованные в формуле символы следует расшифровывать.

Правила оформления графики.

- ◆ Растровые форматы: рисунки и фотографии, сканируемые или подготовленные в Photoshop, Paintbrush, Corel Photopaint, должны иметь разрешение не менее 300 dpi, формата TIF, без LZW уплотнения, CMYK.
- ◆ Векторные форматы: рисунки, выполненные в программе CorelDraw 5.0-11.0, должны иметь толщину линий не менее 0,2 мм, текст в них может быть набран шрифтом Times New Roman или Arial. Не рекомендуется конвертировать графику из CorelDraw в растровые форматы. Встроенные — 300 dpi, формата TIF, без LZW уплотнения, CMYK.

По вопросам публикации следует обращаться к шеф-редактору научно-практического журнала «Современная наука: актуальные проблемы теории и практики» (e-mail: redaktor@nauteh.ru).